Loading analyte results to the Geochemistry Database using the MS ACCESS LOADER application

BRITISH GEOLOGICAL SURVEY

INFORMATION MANAGEMENT PROGRAMME

INTERNAL REPORT IR/06/097

Loading analyte results to the Geochemistry Database using the MS ACCESS LOADER application

A C Mackenzie and C C Johnson

Keywords

Geochemistry, Database, G-BASE.

Keyworth, Nottingham   British Geological Survey   2006

BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham, Edinburgh and London; 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.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

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 its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.

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

British Geological Survey offices

Keyworth, Nottingham NG12 5GG

☎ 0115-936 3241   Fax 0115-936 3488
e-mail: sales@bgs.ac.uk
www.bgs.ac.uk
Shop online at: www.geologyshop.com

Murchison House, West Mains Road, Edinburgh EH9 3LA

☎ 0131-667 1000   Fax 0131-668 2683
e-mail: scotsales@bgs.ac.uk

London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE

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

Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU

☎ 01392-445271   Fax 01392-445371

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

☎ 028-9038 8462   Fax 028-9038 8461

Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB

☎ 01491-838800   Fax 01491-692345

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

☎ 029–2052 1962   Fax 029–2052 1963

Parent Body

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

☎ 01793-411500   Fax 01793-411501
www.nerc.ac.uk

Acknowledgements

The authors thank Dee Flight for reading through this report and contributing useful comments for improving it.

Contents

FIGURES

TABLES

Summary

LOADER is a MS ACCESS application to assist the bulk loading of sample site and analytical data, written specifically for the G-BASE (Geochemical Baseline Survey of the Environment) project but it is also applicable to the bulk loading of any inorganic geochemical sample data. This report does not cover the bulk loading of site information as this is reported elsewhere. It covers the bulk loading of inorganic analytical analyses to Oracle data tables of the British Geological Survey's (BGS) Geochemistry Database.

The process of loading the data consists of three stages. Firstly, the data provider is required to format the data in an MS EXCEL worksheet to a layout specified in this report. Secondly, the data provider submits a request to load or update data to the Geochemistry Database Data Manager. An important part of this process is that the data provider must sign off the data as having been loaded correctly. Finally, the LOADER application is run to load data to the relevant Oracle data tables. This final stage can only be undertaken by Oracle users with 'access' and 'write' permissions so most data providers will not be involved in this stage. Details of the LOADER application are included in the report so the complete process of loading the data can be understood.

Raw data can now be transferred direct into raw data tables of the Geochemistry Database via the BGS Laboratory Information Management System (LIMS) but there is still a requirement to bulk load conditioned data (i.e. data that has undergone a number of QA/QC and levelling procedures) and capture legacy raw and conditioned data not yet entered into the database.

1 Introduction

A Microsoft ACCESS application called "LOADER" to upload geochemical results from the G-BASE project to the corporate Oracle Geochemistry Database has been written by Alan Mackenzie. The Geochemistry Database, its data and domain (code dictionaries) tables, and its history are described by Harris and Coats (1992) and Coats (2004). LOADER is used to load analytical data for a variety of sample types including soils, stream sediments and waters. The application can also be used to load non-G-BASE data but the geochemical results must be presented in the format as described in this report and the dataset must use the Geochemistry Database design as a basis for its sample numbering. LOADER is also used to transfer G-BASE field data to the Geochemistry Database and as this is described in the G-BASE field database manual (Lister et al, 2005) the loading of field data will not be discussed here. This report supersedes the initial documentation written to accompany the LOADER programme (Mackenzie, 2002).

Geochemistry data is usually provided with columns of analytical results for rows of geochemical samples. One of the main functions of LOADER is to transform the data into the structure required by the Geochemistry Database, notably to list each analyte determination with information (such as coded method of analysis) by rows.

Raw or modified ("conditioned") analytical results can be loaded and are stored in different database tables which are summarised in Figure 1.

Raw analytical data for soils and sediments is now directly transferred from the laboratory information management system (LIMS) but older raw G-BASE results and analytical data from other projects still need to be loaded manually to the raw data tables, and conditioned data (i.e. raw data made fit-for-purpose by a series of QA/QC and levelling procedures) also needs to be bulk loaded.

The input data format for the LOADER programme is an MS EXCEL worksheet. More than one worksheet can be used to load analyses from different methods or calibrations, and generally the site information will be loaded before the analytical results. As part of the analytical data submission procedure, the Data Manager will decline to load conditioned data for which there is no spatial reference.

The Excel worksheet should have a header consisting of three rows; row 1 should contain the column names; row 2 analyte lower detection limits; and row three a character for each column that establishes the correct column format (i.e. whether numeric integer or text etc.). There are eight compulsory columns that must be included in the worksheet, (NUMBERING_SYSTEM, PROJECT, SITENO, SAMPLE_TYPE, DUPLICATE, LAB, BATCH_ID and METHOD) and these form the unique key in the MS ACCESS database. From these mandatory fields data can be loaded to the BGS_DTA_BATCHES, BGS.DTA_ANALYTES and BGS_DTA_ANALYSIS data tables (in the case of modified data). Additional useful information (such as date of analyses, and a greater range of analytical limits) can be manually added to the database on request to the Data Manager.

LOADER is an application to append **new** rows of data to the Geochemistry Database and is not used for updating existing records. If a large number of records need to be updated then existing records should be deleted first from the database, an action only the Database Manager and Data Manager are authorised to do.

A sample site has British National Grid coordinates (defined to the nearest metre) and will, when combined with the project/area code, sample type and numbering system code be a unique id. It is also possible to use Irish Grid coordinates to accommodate samples collected from N Ireland. It is worth noting here that G-BASE surface and deep soils (sample types "A" and "S"

respectively) do not conform strictly to the rules of database as sites are given the same number (to make data recording and interpretation simpler), even though more correctly they have differing x, y, z coordinates. In order to conform to the Geochemistry Database design, the LOADER program adds 50,000 to site numbers that are < 50000 if the sample type it is "S" (deep soil).

The design of the LOADER application relies on a series of queries, run within a macro that load data to temporary tables in MS ACCESS and Oracle from where the data can be transferred to the Geochemistry Database after some verification checks and conversion of oxide percentages to ppm are made. LOADER relies on a large number of predefined MS ACCESS queries and this is why it requires the column names in the data to be loaded, to be consistent with those expected by the query.
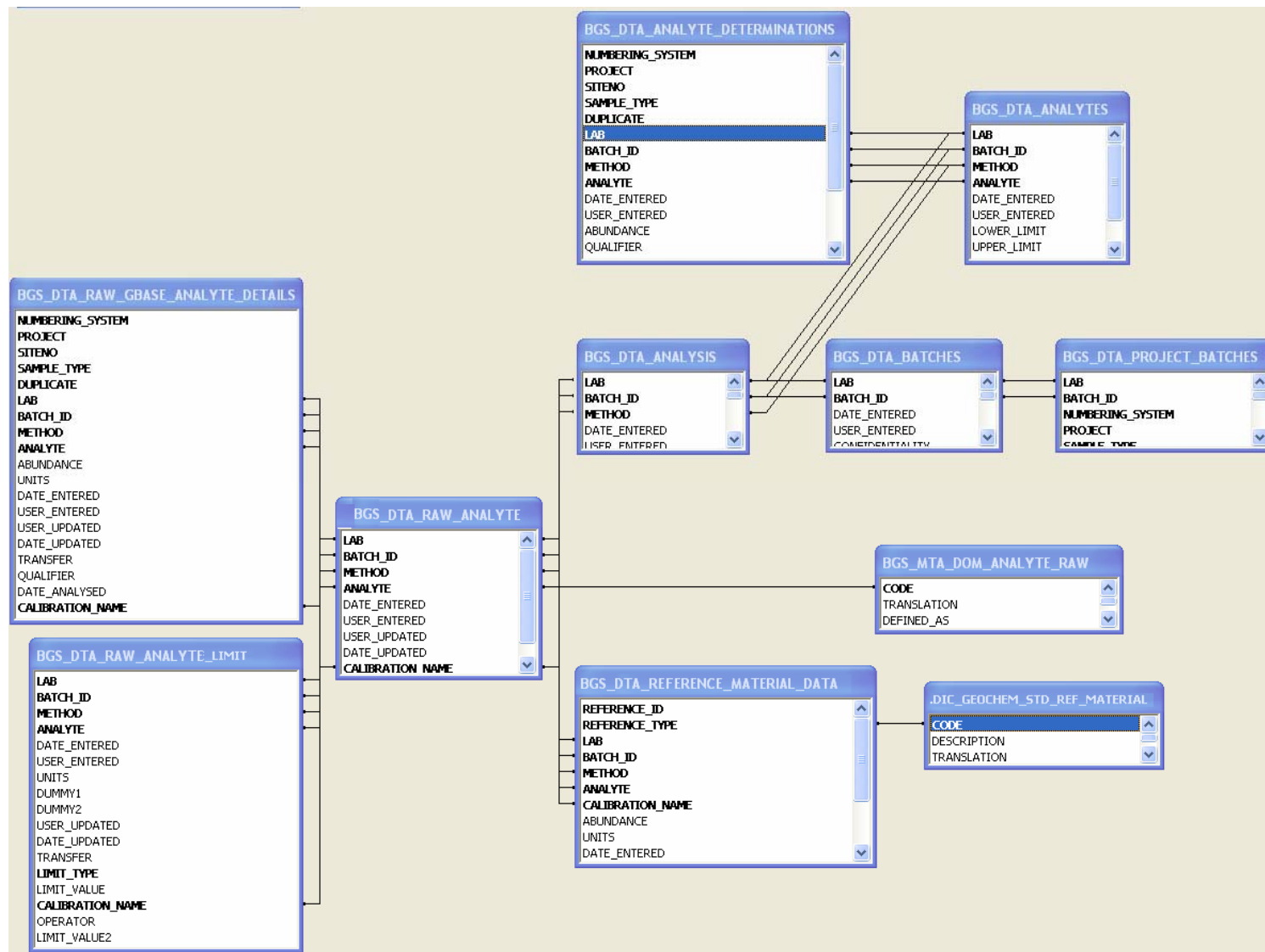
Figure 1: Relationship between tables into which analytical data is loaded

2 The MS ACCESS LOADER Application

2.1 HISTORY OF USE AND DEVELOPMENT

The MS ACCESS LOADER application was written by Alan Mackenzie (Geochemistry Database Manager) in 2002 to load bulk G-BASE sample site and analytical data to the corporate Geochemistry Database. It has been used almost exclusively for the G-BASE project and operated by Alan Mackenzie. The other traditional main data provider to the database was the Mineral Reconnaissance Project (MRP) and the current Geochemistry Database Data Manager (Sue Hobbs) was responsible for directly loading much of this data. However, with more environmental and soil projects generating geochemical data there is a need to capture information from wider sources. In 2006 the Geochemistry Information Management Project (IM4 (Geochemistry) – Project Manager Chris Johnson) improved procedures for the submission of data to the database (see Appendix 1) and this report, to help data providers prepare their data for loading, is part of the procedural strengthening.

An important function of the LOADER application, discussed by Lister et al (2005), is the fact that some of the G-BASE field codes are translated during the loading process from the long established and simple codes used by the G-BASE project to codes that conform to BGS database dictionaries. This applies mainly to the field data, the only conversion of the analytical data is the translation of oxide data reported as % into the required elemental concentrations in ppm (see Section 3.2.2).

For reference, code from the LOADER application is given in Appendix 2 and it is acknowledge that much of this was based on a concept by Andrew McKenzie (BGS Wallingford) developed by Alan Mackenzie in 1998. The complex and numerous macros run by LOADER are best seen by opening the application which is currently located at:

\\kwsan\WorkSpace\IM\CorporateISMaintenance\Code\Geochemdatabase\field_and_analyses\load_geochm.mdb

The last update at the time of writing this report was 17 July 2006 when a greater range of qualifier characters were enabled in the executable code.

2.2 DESCRIPTION OF WHAT THE APPLICATION DOES
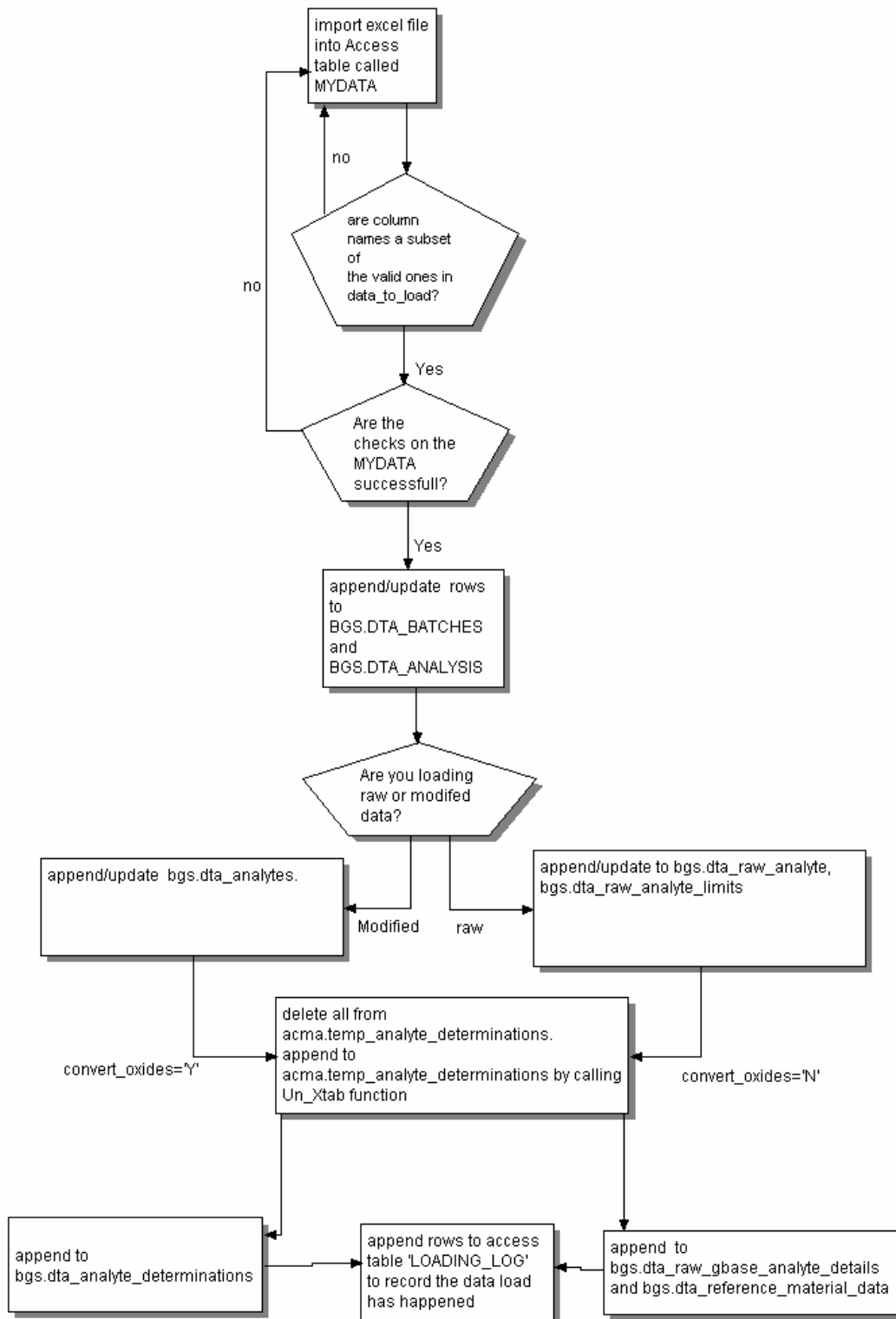
Figure 2: Figure summarising the processes carried out by the LOADER application.

3 Procedure for loading analytical results to the Geochemistry Database

3.1 REQUIREMENTS

The procedure for submitting data to the Geochemistry Database is described in "Geochemistry Database: Procedure for amending or appending data"[1]. This involves submitting a digital file of the data in a predefined format along with a request to the Database Data Manager to load the data. An important part of this process is that the data provider must confirm that the data has been loaded correctly. For most data providers the only concern will be preparing the data for loading to the database and submitting a request to load data. Software requirements shown below are only required if the LOADER application is being used to load data to the database, a process for which only the Database Manager and Data Manager have permissions.

3.1.1 Software Requirements

In order to use the LOADER application there are the following software requirements:

- A PC running MS ACCESS 2000

- A correctly set-up ODBC (**O**pen **D**ata**B**ase **C**onnectivity) connection to the relevant Oracle data source (kk)

- An Oracle user id

- Select permissions on the BGS_MTA_DTA* and insert permissions on BGS_DTA_* oracle tables

- Access to the LOADER application at:

  \\kwsan\WorkSpace\IM\CorporateISMaintenance\Code\Geochemdatabase\field_and_analyses\load_geochm.mdb

  Note that the application can only be used by one person at a time.

3.1.2 Data requirements

- The sample numbering system should correspond to that required by the Geochemistry Database (see Section 3.2.1.1)

- Analytical data should always be accompanied by information about which laboratory did the analyses, the method of analysis and a reference/laboratory batch number (see Section 3.2.1.1)

- The analyte names should correspond to those defined in the database table BGS_MTA_DOM_ANALYTE (see Section 3.2.1.2)

- All analytical data should have corresponding site and sample information such as grid coordinates loaded separately as described in Lister et al (2005). The Data Manager will not allow the loading of analytical results for samples that have no site coordinates.

---

[1] ERMS Object ID = D8408

3.2 PREPARING THE ANALYTICAL DATA FOR LOADING

Preparing the data for loading requires the data provider to prepare an Excel spreadsheet with the data in a specified format and in which the content of the data fields correspond to the requirements of the Geochemistry Database.

3.2.1 Layout of the Excel Spreadsheet

An example of the layout of a data submission is given in Figure 3. There are a number of compulsory columns as described below, but the columns can be in any order and should have one or more analyte column. Similarly there are a number of required header rows and there should be one or more rows of sample analyses. The content of the non-analyte columns requires compliance with Geochemistry Database codes and it is important that planning how samples are to be correctly numbered should be undertaken at the onset of the sampling project. There are ways of dealing with non-compliant data and these should be discussed with the Database Data Manager.

3.2.1.1 COLUMNS

NUMBERING_SYSTEM

A single digit **numeric** code to classify the BGS numbering system used to originally number samples. This was introduced to deal with legacy data so the possibility of duplicate sample ids was removed. This code is defined in the domain table BGS_MTA_DOM_NUMBERING_SYSTEM. The code for the G-BASE numbering system, for example is 4.

CODE	TRANSLATION	DEFINED_AS
1	Min Pet	
2	Bio Strat	
3	MRP	Mineral Reconnaissance Programme
4	GBASE	Geochemical Baseline Survey of the Environment
5	Environmental surveys	

Table 1: BGS_MTA_DOM_NUMBERING_SYSTEM domain table

PROJECT

This a two character **alphanumeric** code that is part of the geochemistry numbering system that identifies the collector or the collection area or the survey on which the samples were collected. The G-BASE project for example has a system of codes to identify the atlas area or urban centre from which samples were collected as defined in Table 7-1 of the G-BASE Field Procedures Manual (Johnson, 2005). The MRP often used the project code to designate the person responsible for the survey, e.g. HB for Howard Bateson.

SITENO

The site number is a unique **integer number** assigned to a site during the course of a single geochemical survey indicated by a unique project code. The number should be >0 and <100000. Site numbering can be problematic if a site is sampled over a period of time or different samples are collected say from a shallow soil auger hole. Strictly speaking if a sample is collected from a different depth the site of sampling is different so a new site number should be assigned. However, the G-BASE project chose to differentiate between top and deep soils collected at a single site by assigning a different sample type to the top (A) and deep (S) soils. A similar complication can arise at monitoring sites when a sample is repeatedly collected over a period if time. For such monitoring it has been established practice to assign a new site number as the uniqueness of a site number cannot be achieved by a differing date/time of sampling.

STEM	PROJ	SITEN	SAMP	DUP	LAB	BATCH_ID	METHO	Na2O	MgO	Al2O3	SiO2	P2O5	K2O	CaO	TiO2	MnO	Fe2O3	Sc	V	Cr	Co	Ba	Ni	Cu	Zn	Ga	Ge
4	60	-1	A	+	BGSL	06947	XRF	0.3	0.3	0.2	0.1	0.05	0.01	0.3	0.02	0.005	0.05	3	3	3	1.5	1	1.3	1.3	1.3	1	0.5
								#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#
4	60	9501	A	+	BGSL	06947	XRF	0.2	#0.15	8.6	62.5	0.25	0.92	0.38	0.729	0.014	3.46	7	60	77	10.2	387	5.4	21.4	22.4	5.1	#0.25
4	60	9502	A	+	BGSL	06947	XRF	0.2	0.5	10.8	53.6	0.60	1.26	5.02	0.547	0.068	8.83	13	161	110	16.3	428	28.1	22.8	107.5	8.0	#0.25
4	60	9503	A	+	BGSL	06947	XRF	0.2	0.5	12.1	52.6	0.47	1.73	0.74	0.646	0.093	10.83	14	180	131	25.8	483	40.0	15.1	94.9	9.3	#0.25
4	60	9504	A	+	BGSL	06947	XRF	0.2	0.5	11.4	53.4	0.54	1.63	1.80	0.606	0.083	10.13	14	166	145	21.1	480	33.6	17.2	100.0	9.1	#0.25
4	60	9505	A	+	BGSL	06947	XRF	0.2	0.9	10.9	45.1	0.30	2.04	32.29	0.422	0.090	6.00	13	74	54	15.6	341	24.6	18.9	64.0	9.8	#0.25
4	60	9506	A	+	BGSL	06947	XRF	0.2	0.8	15.3	56.6	0.30	2.01	0.77	0.885	0.069	8.28	13	120	105	22.5	461	27.7	18.3	78.4	13.2	0.7
4	60	9507	A	+	BGSL	06947	XRF	0.2	0.6	12.6	53.2	0.41	1.73	1.36	0.662	0.109	10.53	14	160	118	24.2	475	32.8	15.8	87.2	10.5	0.8
4	60	9508	A	+	BGSL	06947	XRF	0.2	0.6	9.2	52.3	0.45	1.35	12.48	0.476	0.094	6.90	15	88	66	17.4	470	33.0	32.4	161.8	5.8	#0.25
4	60	9509	A	+	BGSL	06947	XRF	0.2	0.5	11.3	60.4	0.24	1.47	1.73	0.831	0.067	5.42	12	100	84	16.7	392	20.0	15.9	58.8	8.3	0.6
4	60	9511	A	+	BGSL	06947	XRF	0.2	0.8	11.3	53.6	0.38	1.90	9.82	0.545	0.122	6.44	13	99	75	20.7	397	27.0	24.9	99.8	8.8	#0.25
4	60	9512	A	+	BGSL	06947	XRF	0.2	0.5	12.3	51.1	0.70	1.64	0.44	0.726	0.102	13.07	15	202	136	30.1	452	39.9	22.6	140.8	11.4	#0.25
4	60	9513	A	+	BGSL	06947	XRF	0.5	0.7	11.3	59.1	0.26	1.72	0.66	0.676	0.077	6.29	9	91	80	18.1	450	23.5	14.5	62.1	8.3	#0.25
4	60	9514	A	+	BGSL	06947	XRF	0.2	1.0	16.4	56.1	0.22	2.00	0.78	0.910	0.043	6.67	15	118	100	16.6	469	30.5	16.2	71.0	14.0	0.8
4	60	9515	A	+	BGSL	06947	XRF	0.4	0.6	11.7	59.0	0.32	1.63	0.86	0.652	0.051	6.64	10	96	88	13.6	433	23.3	15.9	79.6	9.3	#0.25
4	60	9516	A	+	BGSL	06947	XRF	0.2	1.1	16.9	52.0	0.31	2.34	1.05	0.930	0.056	7.95	16	142	125	18.9	474	35.0	18.0	95.8	15.1	1.0
4	60	9517	A	+	BGSL	06947	XRF	0.4	0.4	10.2	58.0	0.47	1.47	1.38	0.602	0.068	7.45	12	119	113	18.3	457	24.3	15.9	83.9	6.5	#0.25
4	60	9518	A	+	BGSL	06947	XRF	0.2	0.9	14.5	50.5	0.36	1.70	3.24	0.692	0.113	9.57	15	152	115	21.1	449	41.7	26.0	115.9	13.6	1.2
4	60	9519	A	+	BGSL	06947	XRF	0.2	0.8	13.8	56.9	0.37	2.06	0.83	0.804	0.059	7.31	12	110	94	19.2	453	27.0	19.8	84.2	12.3	#0.25
4	60	9520	A	+	BGSL	06947	XRF	0.4	0.4	9.6	60.5	0.37	1.55	0.47	0.508	0.062	6.19	9	84	71	17.3	424	19.3	17.8	63.0	6.8	1.0
4	60	9521	A	+	BGSL	06947	XRF	0.2	0.5	11.5	48.0	1.03	1.59	2.39	0.584	0.159	16.21	19	301	192	38.3	491	63.7	35.2	222.3	9.2	1.0
4	60	9523	A	+	BGSL	06947	XRF	0.2	0.4	10.6	52.9	0.48	1.43	2.93	0.604	0.091	11.43	15	177	133	28.4	418	39.6	13.3	110.4	7.2	#0.25
4	60	9524	A	+	BGSL	06947	XRF	0.2	0.8	12.9	57.0	0.25	1.93	1.97	0.725	0.067	6.66	13	111	88	19.4	444	28.9	23.2	81.2	10.7	#0.25
4	60	9525	A	+	BGSL	06947	XRF	0.2	1.0	14.1	54.0	0.29	2.16	3.82	0.749	0.138	6.88	15	117	89	22.9	444	36.9	27.5	99.4	13.7	0.8
4	60	9526	A	+	BGSL	06947	XRF	0.4	0.7	11.9	57.2	0.35	1.82	2.36	0.687	0.077	6.96	11	108	91	16.3	450	27.8	21.4	81.2	9.6	#0.25
4	60	9527	A	+	BGSL	06947	XRF	0.2	#0.15	8.8	59.8	0.41	1.24	0.89	0.588	0.079	6.89	10	108	80	14.6	402	14.8	15.1	59.0	6.5	#0.25
4	60	9528	A	+	BGSL	06947	XRF	0.2	1.0	17.3	54.8	0.27	2.28	0.99	0.912	0.059	7.95	13	131	115	19.5	473	32.6	19.0	88.8	16.0	0.8
4	60	9529	A	+	BGSL	06947	XRF	0.2	0.5	9.8	57.0	0.51	1.54	2.50	0.522	0.096	8.93	11	114	84	20.2	420	28.0	25.6	113.7	7.6	#0.25
4	60	9530	A	+	BGSL	06947	XRF	0.2	0.5	11.9	59.4	0.36	1.11	1.39	0.861	0.050	4.73	12	92	82	17.0	410	20.9	33.8	91.7	9.3	1.5
4	60	9531	A	+	BGSL	06947	XRF	0.2	0.5	11.9	57.1	0.38	1.39	0.44	0.742	0.044	6.82	12	148	108	15.9	403	21.0	13.2	77.2	9.1	0.8
4	60	9532	A	+	BGSL	06947	XRF	0.2	0.8	14.1	53.7	0.41	1.93	4.90	0.749	0.068	8.20	15	119	106	19.7	462	32.1	29.0	123.2	11.5	#0.25
4	60	9533	A	+	BGSL	06947	XRF	0.4	0.8	13.2	57.5	0.23	1.92	1.06	0.720	0.113	6.53	14	102	85	18.4	474	31.1	23.0	85.1	11.0	0.7
4	60	9534	A	+	BGSL	06947	XRF	0.2	#0.15	9.0	57.0	0.52	1.48	2.37	0.525	0.116	9.45	12	119	96	18.5	451	26.0	13.1	84.2	6.6	#0.25
4	60	9535	A	+	BGSL	06947	XRF	0.2	0.4	10.0	48.7	0.74	1.41	0.64	0.597	0.164	24.19	15	329	190	40.8	433	54.6	21.8	187.7	9.4	#0.25
4	60	9536	A	+	BGSL	06947	XRF	0.2	0.4	11.0	57.8	0.39	1.50	1.06	0.658	0.065	7.73	12	125	106	18.9	426	23.5	13.0	72.1	7.6	#0.25
4	60	9537	A	+	BGSL	06947	XRF	0.2	#0.15	8.9	59.5	0.51	1.40	0.72	0.491	0.048	7.46	10	119	108	21.3	456	25.1	8.5	78.9	5.7	#0.25
4	60	9538	A	+	BGSL	06947	XRF	0.2	0.8	12.7	55.8	0.19	1.96	7.90	0.800	0.064	4.63	12	88	76	13.8	386	20.5	16.4	50.4	9.9	#0.25
4	60	9539	A	+	BGSL	06947	XRF	0.2	0.8	13.9	54.6	0.42	1.91	1.48	0.696	0.064	9.88	14	156	131	21.8	467	40.4	16.5	97.6	13.2	#0.25
4	60	9540	A	+	BGSL	06947	XRF	0.2	0.4	9.7	53.1	0.61	1.53	1.09	0.503	0.142	11.66	14	178	120	22.2	471	37.1	18.1	112.1	7.7	#0.25

Figure 3: Example of format for submitting data as MS EXCEL worksheet (note leading columns have been compressed to save space)

SAMPLE_TYPE

A single character **alphabetic** code that designates the type of sample media being collected. This is defined in the BGS_MTA_SAMPLE_TYPE domain table (Table 2).

CODE	TRANSLATION	DEFINED_AS
A	Shallow G-BASE soil sample	Shallow G-BASE soil sample. Standard depth 5-20cm.
C	Stream Sediment	Stream Sediment
D	Drill core	Drill core.
E	Mineral	Mineral.
F	Float sediment (Fines)	Float sediment (Fines).
G	Gas	Gas.
H	Panned Rock	Panned Rock
J	Marine sediment	Marine sediment.
M	Drill mud (Sludge)	Drill mud (Sludge).
N	Panned sludge	Panned sludge.
P	Panned stream sediment	Panned stream sediment.
R	Rock	Rock.
S	Soil	Shallow overburden altered by soil forming processes. G-BASE sample generally 35-50cm
T	Deep overburden	Deep overburden unaltered by soil forming processes (often a till in northern Britain).
U	Panned overburden or soil	Panned overburden or soil.
V	Vegetation	Vegetation.
W	Water	Water.
Z	Mine Waste	Tailings etc

Table 2: BGS_MTA_SAMPLE_TYPE domain table

DUPLICATE

For the G-BASE project this column is not relevant as the project always numbers duplicates and replicates with a different sample number so as to be undetectable by the laboratories as control samples. For the G-BASE project and most other projects, the entry for this column will simply be a "+". This column can be used to distinguish sites or analytical data that has the same site number (for example, the sample may have been split into different fraction sizes before analysis). By placing a 1-9, or A-Z single character code in this column the uniqueness of the sample id key can be maintained.

LAB

This is a four letter alphanumeric code to designate where the samples where analysed. The BGS laboratories are, for example, BGSL. The code should be defined in the domain table BGS_MTA_DOM_ORGANISATION.

BATCH_ID

The laboratory carrying out the sample analysis should have a unique laboratory reference number to identify one analytical batch of samples. The structure and format of this code is not controllable but is generally a five character alphanumeric code.

METHOD

This is a maximum of six alphanumeric character code to define the method of analysis used to determine the analyte abundance. Codes are defined in the domain table BGS_MTA_ANALYSIS_METHOD.

CODE	TRANSLATION	DEFINED_AS
AAS	Atomic Absorption Spectrometry	Atomic Absorption Spectrometry.
BERYL	Beryllometry	Gamma excitation of Beryllium and detection of neutrons.
BETAP	Beta Probe	Direct electron excitation X-ray spectrometry.
COL	Colorimetric Method	Colorimetric Method
COLHEX	Colorimetric Hot Extractable	Colorimetric Hot Extractable
DNA	Delayed Neutron Activation	Delayed Neutron Activation
DCOES	Direct Reading OES	Direct Reading Optical Emission Spectroscopy
ETAAS	Electrothermal AAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Fire Assay AAS	Fire Assay Atomic Absorption Spectroscopy
FDCP	Fire Assay Direct Coupled Plasma	Fire Assay Direct Coupled Plasma.
FICP	Fire Assay ICP-AES	Fire Assay Inductively Coupled Plasma Atomic Emission Spectrometry
FICPMS	Fire Assay ICP-MS	Fire Assay Inductively Coupled Plasma Mass Spectrometry
FLAAS	Flame AAS	Flame Atomic Absorption Spectrometry
FLAES	Flame Atomic Emission Spectrometry	Flame Atomic Emission Spectrometry
FNA	Fire Assay Neutron Activation	Fire Assay Neutron Activation
GRAV	Gravimetric method	Gravimetric method
HYICP	Hydride Generation ICP	Hydride Generation Inductively Coupled Plasma
ICP	Inductively Coupled Plasma	Inductively Coupled Plasma
ICPMS	ICP Mass Spectrometry	Inductively Coupled Plasma Mass Spectrometry
ISE	Ion Selective Electrode	Ion Selective Electrode
LAICP	Laser Ablation ICP-MS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
NA	Neutron Activation	Neutron Activation
OES	Optical Emission Spectroscopy	Optical Emission Spectroscopy
SPPHOT	Spectrophotometric method	Spectrophotometric method
TITRE	Titrimetric method	Titrimetric method
XRF	X-Ray Fluorescence Pressed Pellets	X-Ray Fluorescence Pressed Pellets
XRFB	X-Ray Fluorescence On Fused Beads	X-Ray Fluorescence On Fused Beads
ICPAES	ICP Atomic Emission Spectrometry	Inductively Coupled Plasma Atomic Emission Spectrometry
LIFS	Laser Induced Fluorescence Spec	Laser Induced Fluorescence Spectroscopy using Scintrex machine
LECO	LECO CHNS Analyser	LECO CHNS Analayser
CE1500	CARLO ERBA 1500 CHNS Analyser	
PE240	PERKIN-ELMER 240 CHNS Analyser	
CVAAS	Cold Vapour AAS	Cold Vapour Atomic Absorption Spectrometry
AQTOC	Shimadzu 5000	
AFS	Atomic Fluorescence	Atomic Fluorescence
NPOC	Organic Content	measurement by non-dispersive infra red of the CO_2 from combusted sample that had been acidified and put on Pt Column at 680c
LOI450	Loss on ignition at 450C	Gravimetric method to determine H_2O content
LOI105	loss on ignition at 1050C	Gravimetric method to determine organic carbon content
XRFWDT	WD X-Ray Fluorescence trace elems	
XRFWDM	WD X-Ray Fluorescence major elems	
XRFEDT	ED-X-Ray Fluorescence trace elems	
XRFWD	WD X-Ray Fluorescence	
XRFED	ED-X-Ray Fluorescence	
XRFS	XRFS	
pH	pH electrode	

Table 3: BGS_MTA_DOM_ANALYSIS_METHOD domain table

<u>Analyte Columns 1 to n</u>

There can be 1 to 247 columns of analyte determinations. Analyte names must match those define in the BGS_MTA_DOM_ANALYTE domain table (Table 4). The exception to this is that any "." present in the analyte name should be replaced with a "_" as the "." causes problems during processing in MS ACCESS.

CODE	TRANSLATION		CODE	TRANSLATION		CODE	TRANSLATION		CODE	TRANSLATION
Ac	Actinium		Dy	Dysprosium		MnO4.-	permanganate		Se.4+	selenite
Ag	Silver		Er	Erbium		MnO4.2-	manganate		Se.6+	selenate
Al	Aluminium		Es	Einsteinium		Mo	Molybdenum		Si	Silicon
Am	Americium		Eu	Europium		N	Nitrogen		Sm	Samarium
Ar	Argon		F	Fluorine		Na	Sodium		Sn	Tin
As	Arsenic		F.-	fluoride		Nb	Niobium		SO3.2-	sulphite
As(III)	As oxidation state 3		Fe	Iron		Nd	Neodymium		SO4.2-	sulphate
As(V)	As oxidation state 5		Fe.2+	ferrous		Ne	Neon		Sr	Strontium
As.3+	arsenite		Fe.3+	ferric		NH3	ammonia		Ta	Tantalum
As.5+	arsenate		Fm	Fermium		NH4.+	ammonium		Tb	Terbium
At	Astatine		Fr	Frankium		Ni	Nickel		Tc	Technetium
Au	Gold		Fulvic	fulvic acids		No	Nobelium		TDS	total dissolved solids
B	Boron		Ga	Gallium		NO2.-	nitrite		Te	Tellurium
B4O7.2-	tetraborate		Gd	Gadolinium		NO3.-	nitrate		Th	Thorium
Ba	Barium		Ge	Germanium		Np	Neptunium		Th.230	thorium - 230
Be	Beryllium		H	Hydrogen		O	Oxygen		Th.232	thorium - 232
Bi	Bismuth		H.3	tritium		OH.-	hydroxide		Ti	Titanium
Bk	Berkelium		H2O.-	moisture		Os	Osmium		TIC	total inorganic carbon
BO2.-	metaborate		H2O.+	combined water		P	Phosphorus		Tl	Thallium
BO3.3-	borate		H2PO4.-	dihydrogen orthophosphate		Pa	Protactinium		Tm	Thulium
Br	Bromine		H2S	hydrogen sulphide		Pb	Lead		TOC	total organic carbon
Br.-	bromide		HCO3.-	bicarbonate		Pd	Palladium		TON	total oxidised nitogen
BrO3.-	bromate		He	Helium		pH	pH		U	Uranium
C	Carbon		Hf	Hafnium		Pm	Promethium		U.234	uranium - 234
C.14	carbon - 14		Hg	Mercury		Po	Polonium		U.238	uranium - 238
Ca	Calcium		Ho	Holmium		P-o	Olsen Phosporous		V	Vanadium
Cd	Cadmium		HPO4.2-	hydrogen orthophosphate		PO4.3-	phosphate		W	Tungsten
Ce	Cerium		Humic	humic acids		Pr	Praseodymium		Xe	Xenon
CEC	Cation exchange capacity		I	Iodine		P-r	Reactive Phosporous		Y	Yittrium
Cf	Californium		I.-	iodide		Pt	Platinum		Yb	Yitterbium
Cl	Chlorine		In	Indium		Pu	Plutonium		Zn	Zinc
Cl.-	chloride		IO3.-	iodate		Ra	Radium		Zr	Zirconium
ClO3.-	chlorate		Ir	Iridium		Ra.226	radium - 226			
Cm	Curium		K	Potassium		Rb	Rubidium			
CN.-	cyanide		Kr	Krypton		Re	Rhenium			
Co	Cobalt		La	Lanthanum		Rh	Rhodium			
CO3.2-	carbonate		Li	Lithium		Rn	Radon			
Cr	Chromium		LOI	loss on ignition		Rn.222	radon - 222			
Cr2O7	dichromate		LOI1000	Loss on ignit at 1000 C		Ru	Ruthenium			
CrO4.2-	chromate		LOI1050	loss on ignition at 1050		S	Sulphur			
Cs	Caesium		LOI450	loss on ignition at 450		S2O3.2-	thiosulphate			
Cu	Copper		Lr	Lawrencium		Sb	Antimony			
d18O	delta oxygen eighteen		Lu	Lutecium		Sc	Scandium			
d2H	delta deuterium		Md	Mendelevium		Se	Selenium			
DIC	Dissolved Inorganic Carbon		Mg	Magnesium		Se(IV)	Se oxidation state four			
DOC	Dissolved Organic Carbon		Mn	Manganese		Se(VI)	Se oxidation state six			

Table 4: BGS_MTA_DOM_ANALYTE domain table

The table BGS_MTA_DOM_ANALYTE_RAW is very similar to Table 4 but oxides such as CaO and K_2O are also included.

3.2.1.2 ROWS

<u>Header Row 1 – Column names</u>

The first eight columns are required data headed with the following column headers:

NUMBERING_SYSTEM   PROJECT SITENO SAMPLE_TYPE DUPLICATE LAB   BATCH_ID METHOD

They do not have to be in this order but it is recommended to follow the order given so it is easier to check formatting is correct. The column headers must be spelt exactly as given though it does not matter whether they are in upper or lower case. Subsequent columns contain the analyte fields named according to the BGS_MTA_DOM_ANALYTE domain table as described previously.

<u>Header Row 2 – Lower Detection Limits</u>

For each of the analyte columns the lower detection limit should be entered as a numeric value in the same units the analyte is reported. If this is left empty then the detection limit is not entered into the BGS_DTA_ANALYTES table. It is understood that for some analytical methods (e.g. XRFED) a different detection limit could be cited for each sample depending on the sample matrix. However, the value sought here is a guideline value to help data users appreciate whether analyte results are close the analytical methods lower limit. This value is also used to generate a value to replace semi-quantitative data, for example, G-BASE replaces values recorded as "below detection" with a value ½ the lower detection limit.

In addition to the detection limit information, the compulsory columns of Row 2 are also populated with data that is not loaded to the database but used to control processes in the executable code. As these compulsory columns are used to create the unique key for the MS ACCESS database, the datasheet cannot contain identical entries in any two rows. In order to distinguish this row from the third row data providers are ask to populate these first eight columns with the same data as that entered in the first eight columns of the fourth row (i.e. the first row of data). The exception is that "-1" should be entered into the SITENO column (see example in Figure 3). This minus value ensures the uniqueness of the record key and is used by the program to confirm that this is the row that contains the lower detection limit information. This is a consequence of a previous data format which allowed the detection limit data to be placed on any row, usually at the end of the data.

<u>Header Row 3</u>

In row 3 a "#" character should be inserted in each of the analyte columns. This is a programming fix to ensure MS ACCESS assigns a text attribute to the data column which is necessary for processing the qualifier code prefix (see Section 3.2.3). If a qualifier code does not appear in the first 20 rows of data then ACCESS assigns a numeric attribute to that field in the database whereas the LOADER program works on the basis of data value being converted to a string rather then numeric value. It is important that the eight compulsory columns contain nothing, i.e."nulls" – a blank space will cause the columns to be a text rather than numeric format.

<u>Rows 4 – onwards</u>

Sample results including compulsory analytical information in first eight columns. There should be at least one record of data and the only limit being the number of rows available in the EXCEL spreadsheet (65,535). If the data has been produced as separate analytical batches or produced by different analytical methods, each data set can be loaded from a separate worksheet. See sections below for comments on the units of measurement, number of decimal places and handling semi-quantitative or absent data. Note that all analyte determinations in a row are

assigned the method of analyses that is shown in the METHOD column of that row. If a different method is to be specified, then the analytical data from the different method should be given on another row or in another worksheet.

3.2.2 Units of Measurement

All analytical data in the database (non-raw data) is stored in parts per million (ppm) except pH, CEC, LOI450 and LOI1050. Data recorded in other units (e.g. ppb) must be converted to ppm. The exception to this is concentrations recorded in oxides for the elements shown in Table 5. These oxides, which must be submitted as percentages, are converted automatically by the LOADER application. The oxide result in percent is multiplied by the conversion factor (Table 5) to give the element concentration in ppm. More rarely reported oxides (such as FeO) will not be identified by the program and must be converted beforehand by the data provider. It is important that to return to the original oxide percentages the same factors are used. In order to return ppm values back to the original oxide percentages, the ppm value should be divided by the factors given in Table 5.

Oxide	Conversion Factor
Al2O3	5290
CaO	7147
Fe2O3	6994
K2O	8301
MgO	6031
MnO	7745
Na20	7419
P2O5	4364
SiO2	4670
TiO2	5995
SO3	4005

Table 5: Oxide percentage to ppm conversion factors used by LOADER.

The number of decimal places permitted is controlled by the numeric format of the ORACLE program i.e. 14 digits are allowed in the number with up to 7 behind the decimal point. If loading raw data then oxide percentages are loaded directly without any need for changing units whilst all other columns of raw data are assumed to be in ppm (mg/kg) units. It is important that data submitted for loading has the same number of decimal places as originally reported by the analyst

3.2.3 Semi-quantitative or incomplete data

Semi-quantitative or incomplete data are represented by qualifier codes defined in the BGS_MTA_DOM_ANALYSIS_QUALIFIER domain table (Table 6). In the past this qualifier has been neglected but is now considered vital information for the data user to be able to assess data quality. The designation of qualified data is the most time consuming part of preparing the data for loading.

The qualifier codes are added as a prefix to the analyte abundance to which they apply. When the data value is loaded to the Geochemistry Database it is the qualifier field of the BGS_DTA_ANALYTE_DETERMINATIONS data table that is populated. LOADER was

originally designed to load processed (i.e. data that has undergone one or more data conditioning processes) rather than raw data (i.e. data as received direct from the laboratory). Since 2005 raw G-BASE raw data tables have been added to the database and when loading legacy raw data the user should replace semi-quantative values (e.g. qualified data) with one of the codes shown in Table 7. Raw G-BASE data for soils and sediments is now loaded directly to the raw data tables via the Laboratory Information Management System (LIMS) (Ault et al, 2006).

Data with no qualifier will have not have an entry in the corresponding analyte qualifier field (i.e. the qualifier field will contain a null entry). Data values that are null *and* which have no qualifier value are not loaded. Some analytical methods will generate data of zero and negative values which may not be removed or modified during the data conditioning process (see Lister and Johnson, 2005).

CODE	TRANSLATION	DEFINED_AS
>	Probably high	Not determined accurately due to interference. Probably higher.
<	Probably low	Not determined accurately due to interference. Probably lower.
^	No estimate possible	Not determined due to interference. No estimate possible.
*	Dubious quality	The value has a documented quality control issue that should restrict its use.
#	Estimated value	Value reported was below the lower detection limit cited by the analyst and has been set to half this detection limit.
$	Uncertain value	Value as reported was below the lower detection limit cited by the analyst.
&	Uncertain quality	No information exists regarding quality of this data.
!	Not available	Applicable, but try as we might, can't find a value.
?	Not entered	No value has been assigned yet (and it might not be applicable).
~	Insufficient sample	Insufficient sample for analysis.
D	Uncertain value	Both $ and *apply to this value.
\	Not applicable	Not requested.
B	Estimated value	Both # and A apply to this value.
A	Estimated value	Value has been set to zero because data conditioning process created an artificial negative value.
C	Uncertain value	Both $ and A apply to this value.

Table 6: BGS_MTA_DOM_ANALYSIS_QUALIFIER domain table

Code	Comment
-94	Insufficient sample (e.g. sample collected but not enough to analyse or lost)
-95	Not determined because of high concentration; but exceeds calibration limit
-96	Not determined because of interference; probably of high concentration
-97	Not determined because of interference; probably of low concentration
-98	Not determined because of interference; no estimate
-99	Absent data (e.g. not requested)

Table 7: Laboratory codes used to replace semi-quantitative values in raw analytical data

3.3 SUBMITTING DATA FOR LOADING

When data is ready for loading to the database, the data provider should follow the procedures indicated in "Geochemistry Database: Procedure for amending or appending data", an indexed and controlled procedure of the Information Management Programme (ERMS reference Object ID = D8408). The form for submitting data loading/updating requests is given in Appendix 1. In the information part of this form it is useful to give some information about the geographical location of the samples being loaded e.g. Weardale. A question regarding geographical location is asked during the running of the LOADER program.

3.4 USING THE LOADER APPLICATION

The requirements for using LOADER and its location on the Keyworth SAN (Site Area Network) are given in Section 3.1.1. This is an application that will be run by a very small number of people so the user interface is not designed in a particularly user-friendly way. On opening the MS ACCESS LOADER application (see section 2.1 for location) the user will be presented with the screen shown in Figure 4. There is a help button from which the user can get basic information on how to proceed. The help information is reproduced in Appendix 3.

Figure 4: Start-up screen for MS LOADER application

Before commencing the loading of the analytical data the user must first load the data from the EXCEL spread sheet (see Figure 3 for example format). This is done using the *file → get external data* menu option and importing data into an ACCESS table in the standard way. No primary key field is set during the import dialogue and the data table must be called "MYDATA". If a MYDATA table already exists, which it will do if LOADER has been used to load data previously, then the existing MYDATA table needs to be renamed. All the options on the start-up screen relate to sample site information (except for confidentiality which relates to batches) and so is not applicable here to the loading of analytical data. Users should be aware that the confidentiality level is set to "4" by default "*Confidential, normal BGS data not released or work in progress*". Levels

of confidentiality are defined in the BGS_MTA_DOM_CONFIDENTIALITY domain table. Any changes to the level of confidentiality code should be made by applying to the Data Manager.

CODE	TRANSLATION	DEFINED_AS
0	Published	Outside published data from BGS or other sources
1	Published	Freely available published BGS data in openfile report
2	Unpublished, retrieval charge	Supplementary information to open file report, not published. Charged at retrieval cost
3	Unpublished, retrieval charge	Supplementary information to open file report, not published. Charged at retrieval cost
4	Confidential	Confidential, normal BGS data not released or work in progress
5	Highly confidential	Highly confidential, Commercial-in-Confidence

Table 8: Table showing the confidentiality codes defined in the BGS_MTA_DOM_CONFIDENTIALITY table

The user will commence loading of the data by clicking the "Load raw/mod GBASE analysis results." button. A complex series of data checking, temporary table creation and query execution takes place as described in Section 2.2. During this time the user will be asked to give the sample type being entered and asked whether the data is raw or modified (i.e. conditioned). Errors will be flagged up as the data is checked and the user has many opportunities to abort the loading before data is appended to the Database. When loading large data sets it would be very unusual for the data to load successfully without any issues. LOADER generates a large number of queries to check the data, some of these are just advisory so the user can be aware of potential data issues but if unchecked could give rise to an Oracle error that would cause the load to fail. Table 9 and Table 10 illustrate two such queries. The first example indicates a list of samples for which the LOADER application has detected that there are sample sites in the database but no analytical data has been provided for these samples. The second example summarises the qualifier field, where found, for each element. Both these queries could advise the user of potential problems in the data lading process but if the user chooses to proceed then issues involving incorrect qualifiers and missing analyses will be ignored and will only be detected when the data provider is asked to check that the data has been loaded correctly.

NUMBER	BGS_DTA	mydata_	mydata_	mydata_S	BGS_DT	BGS_DT	REL_SA	ATLAS	SAMP_ST	EASTING	NORTHING
	C				44	1836		EANG		637690	323140
	C				44	5496		EANG		559265	279576
	C				44	2082		EANG		627350	316800
	C				44	2398		EANG		597590	322120
	C				44	3014		EANG		588930	308220
	C				44	3112		EANG		606110	307690
	C				44	3812		EANG		611040	312890
	C				44	2821		EANG		544480	289780
	C				44	4445		EANG		653079	284268
	C				44	4416		EANG		616777	294260
	C				44	1416		EANG		630450	328480
	C				44	1452		EANG		633970	328060
	C				44	2541		EANG		546580	305300
	C				44	2844		EANG		547620	293690
	C				44	3028		EANG		563630	289110
	C				44	3037		EANG		587240	307290
	C				44	3040		EANG		587970	307470
	C				44	3115		EANG		595880	306080
	C				44	5619		EANG		558362	272580
	C				44	334		EANG			
	C				44	3092		EANG		568790	289300

Table 9: Results from the query A_C_QRY_CHECK_IF_SED_SITES_HAVE_MISSING_DATA

SAMPLE_	ANALYTE	QUALIFIER	min_abundance	max_abundance	CountOfBATCH_ID	MinOfQUALIFIER
C	Al	A	0	0	13	A
C	Al		0	135953	2333	
C	Ba	#	50	50	9	#
C	Ba	*	100	1500	2336	*
C	Ba		100	100	1	
C	Ca	$	0	1500.87	3	$
C	Ca	C	0	0	1	C
C	Ca		2144.1	2144.1	1	
C	Ca	*	1746.36945	464634.2459	2341	*
C	Cl	*	500	46600	354	*
C	Cl		500	500	1	
C	Cl	#	250	250	413	#
C	Cl	$	100	400	1578	$
C	Fe	*	3024.611252	372291.5712	2345	*
C	Fe		349.7	349.7	1	
C	Mg	B	0	0	6	B
C	Mg	C	0	0	8	C
C	Mg	$	120.62	1085.58	61	$
C	Mg		784.03	24727.1	2271	
C	Na	#	1780.56	1780.56	40	#
C	Na		2225.7	73596.48	1323	
C	Na	$	1187.04	2299.89	983	$
C	P		43.64	33821	2340	
C	P	C	0	0	2	C
C	P	$	0	43.64	4	$
C	S		2002.5	2002.5	1	
C	S	$	400.5	1602	1071	$
C	S	#	1001.25	1001.25	103	#
C	S	*	2002.5	76095	1171	*
C	Si	A	0	0	1	A
C	Si		467	466346.2	2345	
C	Ti		119.9	119.9	1	
C	Ti	$	89.925	89.925	1	$
C	Ti	*	17.985	14049.20876	2342	*
C	Ti	A	0	0	2	A

Table 10: Results from the query A_C_check_temp_analyte_determinations

The person responsible for loading the data should provide, where appropriate, feedback to the data provider concerning "errors" detected during loading. This should include information produced by the "check queries" described previously. Any problems occurring during the append or update operation may result in some or no data being loaded. However, it is unlikely that errors will occur during transfer to the main database because of all the prior checking of the data provided. Once all data is loaded an entry is made in the *loading_log* ACCESS data table of how many rows were loaded to the BGS_DTA_ANALYTE_DETERMINATIONS table for each element.

Appendix 1 : Data submission form

REQUEST TO APPEND/AMEND DATA TO GEOCHEMISTRY DATABASE

Ref. No.	Data Submitted by:	
	Project:	
	Project Code:	

Description of data to be appended/amended:

Instructions:

<This forms is available from ERMS, Object id ref = D8411>

ERMS file reference to detailed instructions:			
	Date:	**Name:**	**Signature:**
Data Provided Submitted:			
Logged by Data Manager:			
Input to Database:			
Data verified by provider:			

Shaded columns will be completed by the Data Manager. Instructions on how to complete this form are given overleaf

Instructions for completing form:

Data Submitted by: This is the name of the person submitting the data e.g. *T.Smith*

Project: Enter the name of the project providing the data e.g. *G-BASE.*

Project Code: BGS project code for the project providing the data e.g. *E1234S99*

Description of data to be appended/amended: This is a brief description of the data that is being edited (amended) or new data being added to the Database. It should be sufficiently detailed for someone in the future and unconnected with a specific project to be able to identify what was being added or amended. Examples would be:

"*New soil analyses from the G-BASE 2005 field campaign in Essex*"

"*Corrections to the Great Glen Atlas area G-BASE data identified by the on-going audit*"

"*Old soil geochemical data (Cu, Pb, Zn and Ag results) from the MRP work on Tow Scar near Ingleton not previously loaded to the Database*"

"*Ag, Bi, Cr, Co, Cu, Mn, Ni, Pb, Sb, Sr, Ti, Zn and Zr results for surface soils analysed as part of the Geochemistry and Health Project (CBH Programme) collected from Mid-Wales by Seán Quigley and Barbara Polumbo in March 2006*"

Instructions: This is where the data provider instructs the Data Manager or Database Manager (the two persons who have edit and write permissions to the Oracle data tables) what to do with the provided data. There are many different scenarios for what might be entered here but the most common will be the addition of new site and/or analytical results. Data providers must be aware of required field, codes that are to be used and the format required for loading. It is not the job of the Data or Database Manager to format data for loading but they will give advice on what is required. There is a protocol to load G-BASE site and analytical data and it is recommended that data providers follow this protocol so data can be automatically loaded using Alan Mackenzie's Microsoft Access "LOADER" application. LOADER cannot be used to update existing results.

ERMS file reference to detailed instructions: If detailed instructions are provided as a separate file then that file should be stored in the BGS ERMS and the document reference number entered here. The Data Manager will create a hardcopy record of all instructions and attach them to the request form.

Signatures: The Data Manager will record when each stage of the loading process has been completed and make sure each task is signed off by the person responsible for carrying out the task. The conclusion of adding new data or the amendment of existing data will be when the data provider has verified that all changes have been carried out satisfactorily as instructed.

Appendix 2 : LOADER code

Function Un_XTab(In_Tab, Pivot, Out_tab, convert_oxides)

```
' function to perform the reverse of an access xtab query

' works by retrieving a list of column names, then running a

' sequence of SQL append queries to build up an output table

'In_Tab is the name of the table to process

'Pivot is the field against which the individual data entries will be written

'Out_tab is the table that the results will be written to

'Out_tab requires three fields:

' ID as Text

' Cname as Text

' Cvalue as text


'original concept, Andrew McKenzie, HG wallingford. Modified by ACMA 1998

' This allows column names to be exactly the same as used by the geochem database. Those column names loaded from the lab files into

'access with e.g. F.- becomes F#- as access can't support the dot. this prog however puts the dot back in again

Dim MyDB As Database

Dim ColName As Object

Dim MyTabledef As Object

Dim querypart1 As String

Dim querypart2, querypart2a As String

Dim querypart3 As String

Dim querypart4 As String

Dim fullquery, fullquery_qual As String

Dim style As String

Dim Msg As String

Dim response As String

Dim anal As String

Dim NUMBERINGSYSTEM As Integer

Dim conv, units As String

Dim name As String

Dim myname As String

Set MyDB = CurrentDb()  ' set current database

Set MyTabledef = MyDB.TableDefs(In_Tab) ' set table definition collection to input table

 NUMBERINGSYSTEM = [Forms]![gbase_loading]![numbering_system]

On Error GoTo UNDO_CHANGES

For Each ColName In MyTabledef.Fields ' iterate through all fields in table
```

```vba
'Debug.Print ColName.name
  myname = UCase(Mid(ColName.name, 1, 1)) & LCase(Mid(ColName.name, 2, 20))
    If Len(ColName.name) > 2 Then
      myname = ColName.name
    End If


If UCase(ColName.name) = "NUMBERING_SYSTEM" Or UCase(ColName.name) = "PROJECT" Or _
   UCase(ColName.name) = "SITENO" Or UCase(ColName.name) = "SAMPLE_TYPE" Or _
   UCase(ColName.name) = "DUPLICATE" Or UCase(ColName.name) = "LAB" Or _
   ColName.name = "BATCH_ID" Or UCase(ColName.name) = "METHOD" Or _
   UCase(ColName.name) = "DATE_ENTERED" Or _
   UCase(ColName.name) = "DATE_ANALYSED" Or _
   UCase(ColName.name) = "CALIBRATION_NAME" Or _
   UCase(ColName.name) = "USER_ENTERED" Then
    'Debug.Print ColName.name
Else
'but only create query below when field name is that of an element
'Debug.Print "i'm in!"
 units = "mg/kg"
  conv = "1.0"


If ColName.name = "CaO" Then
 If convert_oxides = "Y" Then
   conv = "7147.0"
   myname = "Ca"
 Else
   conv = "1.0"
   units = "%"
 End If


ElseIf ColName.name = "Al2O3" Then
 If convert_oxides = "Y" Then
   myname = "Al"
   conv = "5290.0"
  Else
   conv = "1.0"
   units = "%"
 End If
ElseIf ColName.name = "SO3" Then
 If convert_oxides = "Y" Then
   myname = "S"
   conv = "4005.0"
```

```
    Else
      conv = "1.0"
      units = "%"
    End If


ElseIf ColName.name = "Fe2O3" Then


If convert_oxides = "Y" Then
      myname = "Fe"
       conv = "6994.0"
    Else
       units = "%"
       conv = "1.0"
    End If
ElseIf ColName.name = "K2O" Then
If convert_oxides = "Y" Then
myname = "K"
conv = "8301.0"
     Else
       units = "%"
       conv = "1.0"
    End If
ElseIf ColName.name = "MgO" Then
If convert_oxides = "Y" Then
myname = "Mg"
conv = "6031.0"
Else
       units = "%"
       conv = "1.0"
    End If
ElseIf ColName.name = "SiO2" Then
If convert_oxides = "Y" Then
myname = "Si"
conv = "4670.0"
Else
       units = "%"
       conv = "1.0"
    End If
ElseIf ColName.name = "TiO2" Then
If convert_oxides = "Y" Then
myname = "Ti"
conv = "5995.0"
```

```vb
Else
  units = "%"
   conv = "1.0"
 End If
ElseIf ColName.name = "MnO" Then
If convert_oxides = "Y" Then
myname = "Mn"
conv = "7745.0"
Else
  units = "%"
   conv = "1.0"
 End If
ElseIf ColName.name = "Na2O" Then
If convert_oxides = "Y" Then
myname = "Na"
conv = "7419"
Else
  units = "%"
   conv = "1.0"
 End If
ElseIf ColName.name = "P2O5" Then
If convert_oxides = "Y" Then
myname = "P"
conv = "4364.0"
Else
  units = "%"
   conv = "1.0"
 End If

Else
conv = "1.0"
End If

If (InStr(1, myname, "_") = 0) Then
anal = myname
Else
anal = Mid(myname, 1, InStr(1, myname, "_") - 1) & "." & Mid(myname, InStr(1, myname, "_") + 1, 3)
End If
Debug.Print anal
querypart1 = "INSERT INTO " & Out_tab & " (NUMBERING_SYSTEM,PROJECT,SITENO,SAMPLE_TYPE,UNITS,DUPLICATE,LAB,BATCH_ID,METHOD,DATE_ENTERED,USER_ENTERED,QUALIFIER,ANALYTE,ABUNDANCE) SELECT DISTINCTROW "
```

querypart2 = querypart1 & NUMBERINGSYSTEM & ",PROJECT,SITENO,SAMPLE_TYPE,'" & units & "',DUPLICATE,LAB,BATCH_ID,METHOD,date(),'ACMA',null ,'" & anal & "'"

querypart3 = querypart2 & " , cdbl(" & In_Tab & ".["

querypart4 = querypart3 & ColName.name & "])*" & conv & " FROM " & In_Tab & " WHERE siteno is not null and (((mid(" & In_Tab & ".["

'fullquery = querypart3 & ColName.name & "])*" & conv & " FROM " & In_Tab & ";"

fullquery = querypart4 & ColName.name & "],1,1)) in ('0','1','2','3','4','5','6','7','8','9','-')));"

Debug.Print fullquery

DoCmd.SetWarnings False ' Turn off warnings

DoCmd.RunSQL fullquery   ' Run query.

querypart1 = "INSERT INTO " & Out_tab & " (NUMBERING_SYSTEM,PROJECT,SITENO,SAMPLE_TYPE,UNITS,DUPLICATE,LAB,BATCH_ID,METHOD,DATE_ENTERED,USER_ENTERED,QUALIFIER,ANALYTE,ABUNDANCE) SELECT DISTINCTROW "

querypart2 = querypart1 & NUMBERINGSYSTEM & ",PROJECT,SITENO,SAMPLE_TYPE,'" & units & "',DUPLICATE,LAB,BATCH_ID,METHOD,date(),'ACMA',"

querypart2a = querypart2 & "(mid(" & In_Tab & ".[" & ColName.name & "],1,1)) ,'" & anal & "'"

querypart3 = querypart2a

querypart4 = querypart3 & ",cdbl(mid(" & In_Tab & ".[" & ColName.name & "],2,15))*" & conv & " FROM " & In_Tab & " WHERE siteno is not null and (((mid(" & In_Tab & ".["

fullquery_qual = querypart4 & ColName.name & "],1,1)) not in ('0','1','2','3','4','5','6','7','8','9','-')));"

Debug.Print fullquery_qual

DoCmd.RunSQL fullquery_qual   ' Run query.

DoCmd.SetWarnings True ' turn back on warnings

End If

Next ColName

DoCmd.Hourglass False

Exit Function

UNDO_CHANGES:

    MsgBox ("An error occured. You will need to correct the errors indicated in the oracle message and reload all the data. ")

    MyDB.Close

'MsgBox ("Undo operation complete.")

End Function

Public Function Continue_yes_no(question As String)

Dim ans, style

style = vbYesNo

ans = MsgBox(question, style, "", "", 1000)

If (ans = vbYes) Then

Continue_yes_no = True

Else

Continue_yes_no = False

End If

End Function

Public Function validate_data_to_load()

```
'returns a count of the number of rows that
Dim MyDB As Database
Dim query_rows1 As DAO.Recordset
Dim query_rows2 As DAO.Recordset
Dim query_rows3 As DAO.Recordset


Set MyDB = CurrentDb()


Set query_rows2 = MyDB.OpenRecordset("A_C_invalid_method_codes")
Set query_rows1 = MyDB.OpenRecordset("A_C_invalid_sample_type_codes")
Set query_rows3 = MyDB.OpenRecordset("A_C_invalid_lab_codes")


 If (query_rows1.RecordCount = 0 Or query_rows2.RecordCount = 0 Or query_rows3.RecordCount = 0)
Then
 validate_data_to_load = True
 Else
validate_data_to_load = False
End If
If (query_rows1.RecordCount <> 0) Then
DoCmd.OpenQuery ("invalid_sample_type_codes")
End If
If (query_rows2.RecordCount <> 0) Then
DoCmd.OpenQuery ("invalid_method_codes")
End If
If (query_rows3.RecordCount <> 0) Then
DoCmd.OpenQuery ("invalid_lab_codes")
End If
```


End Function

Public Function compare_database(query_name As String)

```
'returns a count of the number of rows that
'Dim MyDB As Object, query_rows2, query_rows3
'Dim query_rows1 As Recordset
'Dim rs As New ADODB.Recordset
Dim MyDB As Database
Dim query_rows1 As DAO.Recordset
Dim query_rows2 As DAO.Recordset
Dim query_rows3 As DAO.Recordset
```

```
Set MyDB = CurrentDb()

Set query_rows1 = MyDB.OpenRecordset(query_name)

 If (query_rows1.RecordCount = 0) Then
 compare_database = False
 MsgBox ("no errors")
 Else
 compare_database = True
 MsgBox ("Errors occurred. The offending rows will be displayed. Query name indicates type of error")
 DoCmd.OpenQuery (query_name)
End If
End Function


Public Function Open_valid_code_table(name As String)
DoCmd.OpenTable (name)
End Function


Public Function Append_to_geochm()
   Dim strName As String
   Dim mess As String
   Dim query_failed As String
   Dim strMessage As String
   Dim wrkDefault As Workspace
   Dim dbsgeochm As Database
     ' Get default Workspace.
   Set wrkDefault = DBEngine.Workspaces(0)
   Set dbsgeochm = CurrentDb()

   On Error GoTo UNDO_CHANGES
 ' Start of outer transaction.
  ' wrkDefault.BeginTrans
   ' Start of main transaction.
   MsgBox ("Starting to load data to the Geochemistry database....")
   DoCmd.OpenQuery ("Append_temp_ds_to_database")
   DoCmd.OpenQuery ("Append_temp_ns_to_database")
   DoCmd.OpenQuery ("Append_temp_drifts_to_database")
   DoCmd.OpenQuery ("Append_temp_clasts_to_database")
   DoCmd.OpenQuery ("Append_temp_sed_ppt_to_database")
   DoCmd.OpenQuery ("Append_temp_geol_to_database")
   DoCmd.OpenQuery ("Append_temp_land_to_database")
```

```
DoCmd.OpenQuery ("Append_temp_sed_to_database")
DoCmd.OpenQuery ("Append_temp_over_to_database")
DoCmd.OpenQuery ("Append_temp_water_to_database")
DoCmd.OpenQuery ("Append_temp_pan_to_database")
DoCmd.OpenQuery ("Append_temp_hor_to_database")
DoCmd.OpenQuery ("Append_temp_min_to_database")
DoCmd.OpenQuery ("Append_temp_cont_to_database")
DoCmd.OpenQuery ("Append_temp_site_cont_to_database")
DoCmd.OpenQuery ("append_temp_clasts_min_TO_database")
DoCmd.OpenQuery ("append_temp_bed_min_TO_database")
    query_failed = "Append_temp_ds_to_log"
DoCmd.OpenQuery ("Append_temp_ds_to_log")
query_failed = "Append_temp_ns_to_log"
DoCmd.OpenQuery ("Append_temp_ns_to_log")
query_failed = "Append_temp_drifts_to_log"
DoCmd.OpenQuery ("Append_temp_drifts_to_log")
query_failed = "Append_temp_clasts_to_log"
DoCmd.OpenQuery ("Append_temp_clasts_to_log")
query_failed = "Append_temp_sed_ppt_to_log"
DoCmd.OpenQuery ("Append_temp_sed_ppt_to_log")
query_failed = "Append_temp_geol_to_log"
DoCmd.OpenQuery ("Append_temp_geol_to_log")
query_failed = "Append_temp_land_to_log"
DoCmd.OpenQuery ("Append_temp_land_to_log")
query_failed = "Append_temp_sed_to_log"
DoCmd.OpenQuery ("Append_temp_sed_to_log")
query_failed = "Append_temp_over_to_log"
DoCmd.OpenQuery ("Append_temp_over_to_log")
query_failed = "Append_temp_wat_to_log"
DoCmd.OpenQuery ("Append_temp_wat_to_log")
query_failed = "Append_temp_pan_to_log"
DoCmd.OpenQuery ("Append_temp_pan_to_log")
query_failed = "Append_temp_hor_to_log"
DoCmd.OpenQuery ("Append_temp_hor_to_log")
query_failed = "Append_temp_min_to_log"
DoCmd.OpenQuery ("Append_temp_min_to_log")
query_failed = "Append_temp_cont_to_log"
DoCmd.OpenQuery ("Append_temp_cont_to_log")
query_failed = "Append_temp_site_cont_to_log"
DoCmd.OpenQuery ("Append_temp_site_cont_to_log")
query_failed = "append_temp_bed_min_TO_log"
DoCmd.OpenQuery ("append_temp_bed_min_TO_log")
```

```
    query_failed = "append_temp_clasts_min_TO_log"
    DoCmd.OpenQuery ("append_temp_clasts_min_TO_log")


MsgBox ("Successfully completed loading data to the Geochemistry database.")
        ' Ask if the user wants to commit to all the changes
        ' made above.
        'If MsgBox("Save all changes?", vbYesNo) = vbYes Then
         '   wrkDefault.CommitTrans
        'Else
        '    wrkDefault.Rollback
       ' End If
           dbsgeochm.Close
'        wrkDefault.Rollback
        'wrkDefault.Close
Exit Function
UNDO_CHANGES:
MsgBox ("An error occurred. . Reload the data once you have corrected the problem, to overwrite what
has been partially loaded")
    mess = "the query that failed was " & query_failed
    MsgBox (mess)
    DoCmd.OpenQuery ("del_temp_hor_to_database")
    DoCmd.OpenQuery ("del_temp_min_to_database")
    DoCmd.OpenQuery ("del_temp_cont_to_database")
    DoCmd.OpenQuery ("del_temp_sed_to_database")
    DoCmd.OpenQuery ("del_temp_over_to_database")
    DoCmd.OpenQuery ("del_temp_water_to_database")
    DoCmd.OpenQuery ("del_temp_pan_to_database")
    DoCmd.OpenQuery ("del_temp_drifts_to_database")
    DoCmd.OpenQuery ("del_temp_clasts_to_database")
    DoCmd.OpenQuery ("del_temp_sed_ppt_to_database")
    DoCmd.OpenQuery ("del_temp_geol_to_database")
    DoCmd.OpenQuery ("del_temp_land_to_database")
    DoCmd.OpenQuery ("del_ds1")
    DoCmd.OpenQuery ("del_ns1")
    dbsgeochm.Close
MsgBox ("Undo operation complete.")
End Function
Public Function compare_database_type1(query_name As String)
'returns a count of the number of rows that
'Dim MyDB As Object, query_rows2, query_rows3
'Dim query_rows1 As Recordset
```

```vb
'Dim rs As New ADODB.Recordset

Dim MyDB As Database

Dim query_rows1 As DAO.Recordset

Dim query_rows2 As DAO.Recordset

Dim query_rows3 As DAO.Recordset

Set MyDB = CurrentDb()

Set query_rows1 = MyDB.OpenRecordset(query_name)

 If (query_rows1.RecordCount > 0) Then

 compare_database_type1 = True

 MsgBox (" no errors")

 Else

 compare_database_type1 = False

 MsgBox ("Errors occurred. either no detection limits and qualifiers present, or there are duplicate
sitenos")

 DoCmd.OpenQuery (query_name)

End If

End Function
```

=========

```vb
Option Compare Database


'----------------------------------------------------------

' append_analyses1

'

'----------------------------------------------------------

Function append_analyses1()
On Error GoTo append_analyses1_Err


   DoCmd.SetWarnings True

   ' Program to load  GBASE analyses into the Geochemistry database. See 'Help' for information about
the data format expected. Your data needs  to have been  imported and renamed  to be  called
'MYDATA'. Exit program if not sure, when next asked

   Beep

   MsgBox "Program to load  GBASE analyses into the Geochemistry database. See 'Help' for
information about the data format expected. Your data needs  to have been  imported and renamed  to be
called 'MYDATA'. Exit program if not sure, when next asked", vbOKOnly, ""

   ' Unless you specify you are loading raw data later, ALL data, except for major oxides must be in PPM
!!!! No checks are made during loading

   Beep

   MsgBox " Unless you specify you are loading raw data later, ALL data, except for major oxides must be
in PPM !!!! No checks are made during loading", vbOKOnly, ""

   ' To import a table choose FILE---> get external data . navigate to your data file. When prompted for
table to import to , choose  existing table called MYDATA.

   Beep
```

MsgBox "To import a table choose FILE---> get external data . navigate to your data file. When prompted for table to import to , choose existing table called MYDATA.", vbOKOnly, ""

If (Continue_yes_no("Do you want to stop the program?") = True) Then

' do you want to continue. Maybe your data is not ready yet

Exit Function

End If

' delete from data_to_load table (into which table MYDATA is loaded further on)

DoCmd.OpenQuery "A_DELETE_FROM_DATA_TO_LOAD", acNormal, acEdit

If (compare_database_type1("A_C_check_DET_limits_present") = False Or compare_database("A_C_duplicate_sitenumbers") = True) Then

' can't continue if no row of detection limits at the top or if duplicate siteno's

Beep

MsgBox "Cannot continue. There is no row of detection limits. Such a row would have a siteno = -1. Also, failure will occur later in load if no qualifiers have been added.", vbOKOnly, ""

' stops if no row of detection limits on top row

Exit Function

End If

' will now append below your data to a temporary table.The error message will guide you to the column that is wrongly named.....also will always get 1 row error you can ignore for the # row , which isn't used in this table other than to force data type

Beep

MsgBox "Now appending your data to a temporary table.The error message will guide you to the column that is wrongly named.....also will always get 1 row error you can ignore for the # row , which isn't used in this table other than to force data type", vbOKOnly, ""

' ...appending mydata table to data_to_load. this is purely to check that column names are correct. We will ultimately be still loading to the database from MYDATA

DoCmd.OpenQuery "A_append_my_data_to_data_to_load", acNormal, acEdit

' warns if some soil sites in the dta_normal_sites table (from the field database) are not represented in this results data we are loading. best to stop and see why rather than continue.

DoCmd.OpenQuery "A_C_QRY_CHECK_IF_NS_SITES_HAVE_MISSING_DATA", acNormal, acEdit

' Now checking Sample_type, method and lab field values. Rows will be displayed if they have invalid values and the program will stop...

Beep

MsgBox "Now checking Sample_type, method and lab field values. Rows will be displayed if they have invalid values and the program will stop...", vbOKOnly, ""

' warns if some seds sites in the dta_drainage_sites table (from the field database) are not represented in this results data we are loading. best to stop and see why rather than continue.

DoCmd.OpenQuery "A_C_QRY_CHECK_IF_SED_SITES_HAVE_MISSING_DATA", acNormal, acEdit

If (validate_data_to_load() = False) Then

' this stops the macro on next line if there are invalid method,sample_type or lab codes

Beep

MsgBox "The rows displayed have invalid values. Please change them and start loading proecedure again.", vbOKOnly, ""

' end

Exit Function

End If

' Now starting loading your data into a temporary oracle table....(acma.TEMP_ANALYTE_DETERMINATIONS)

Beep

MsgBox "..also, no invalid method,sample_type or lab values", vbOKOnly, ""

DoCmd.SetWarnings False

' deleting from temp_analyte_determinations table to allow data from mydata to be transferred

DoCmd.OpenQuery "A_del_from_temp_analyte_determinations", acNormal, acEdit

' Now starting loading your data into a temporary table....(TEMP_ANALYTE_DETERMINATIONS

Beep

MsgBox " Now starting loading your data into a temporary table....(TEMP_ANALYTE_DETERMINATIONS", vbOKOnly, ""

If (Eval("Continue_yes_no(""Do you want tocontinue. You must if your data is not either % for oxides and mg/kg for the rest. Once changed, you need to run query called ...append temp to raw AD... and adjust units in raw_analytes table?"")=False")) Then

    ' do you want to continue. Maybe you've remembered your data is not ready yet or some field sites do not have analyses and need investigating

    Exit Function

End If

If (Continue_yes_no("are you loading raw data?") = True) Then

    ' if raw data then transferring data from mydata to temp_analyte_determination table. No conversion of Oxides will be done.

    Call Un_XTab('MYDATA','xxx',' TEMP_ANALYTE_DETERMINATIONS','N')

    '   data now transferred into temp_analyte_determinations

    Beep

    MsgBox "your data is now in the temp_analyte_determations table.", vbOKOnly, ""

    '   show summary of temp_analyte_determination so we can see if the ranges are wrong

    DoCmd.OpenQuery "A_C_check_temp_analyte_determinations", acNormal, acEdit

    '   appends minimal batch info to dta_batches. It appends a new row if the batch is not in the batches table. No update of exisiting rows

    DoCmd.OpenQuery "A_C_check_batches_loaded", acNormal, acEdit

    '   appends minimal analyses info to dta_analyses. It appends a new row if the analysis is not in the analyses table. No update of exisiting rows

    DoCmd.OpenQuery "A_C_check_analyses_loaded", acNormal, acEdit

End If

'   appends rows to dta_raw_analyte where there are no existing rows

DoCmd.OpenQuery "A_C_check_raw_analyte_loaded", acNormal, acEdit

'   updates any existing raw detection limits in dta_raw_analyte_limit, from data in the temp_analyte_determination table where not null. Will have not been converted to ppm if it was a major oxide

DoCmd.OpenQuery "A_append temp raw DL to AD", acNormal, acEdit

'   appends any missing detection limits rows in dta_raw_analyte_limit by giving them a null detection limit value

DoCmd.OpenQuery "A_C_check_raw_limits_loaded", acNormal, acEdit

'   appends temp_analyte_determination table to the raw table (bgs.dta_raw_gbase_analyte_details)

DoCmd.OpenQuery "A_append temp to raw AD", acNormal, acEdit

'   inserts row in log file

DoCmd.OpenQuery "A_append temp to raw AD_log", acNormal, acEdit

' macro stops. if continuing below its since we said we are dealing with modified data and so the code below runs instead of that above

Exit Function

' we are here because just dealing with modified data

Beep

MsgBox "just dealing with modified data", vbOKOnly, ""

' this adds 50k to S sitenos where < 50000 . this didnt apply to raw data above

DoCmd.OpenQuery "A_update_my_data_to_data_to_load", acNormal, acEdit

' as we said its not raw data then transferring data from data_to_load to temp_analyte_determination table. conversion of Oxides will be done.

Call Un_XTab('MYDATA','xxx',' TEMP_ANALYTE_DETERMINATIONS','Y')

' summary of data just loaded

Beep

MsgBox "Next is a summary of data just loaded to temporary table", vbOKOnly, ""

' show summary of temp_analyte_determination so we can see if the ranges are wrong

DoCmd.OpenQuery "A_C_check_temp_analyte_determinations", acNormal, acEdit

' below starts process of loading to   bgs.dta_analyte_determination database

Beep

MsgBox "about to append to main database. ....You can stop the append query by choosing cancel when the operation is being undertaken", vbOKOnly, ""

If (Continue_yes_no("Do you want to carry on? if the ranges are not right do not continue. ") = False) Then

    ' do you want to continue with loading to main database in light of any errors thrown up above

    Exit Function

End If

' this and queries below add data to the batches,analyses and analytes table

DoCmd.OpenQuery "A_C_check_batches_loaded", acNormal, acEdit

' adds rows to analyses tables where they do not already exist

DoCmd.OpenQuery "A_C_check_analyses_loaded", acNormal, acEdit

' updates existing Detection Limits in the dta_analytes table if the limit value in our load file is not null

DoCmd.OpenQuery "A_append temp DL to AD", acNormal, acEdit

' appends Detection Limits (even null values) from the temp_analyte_determination table where there are no rows in the dta_analytes table

DoCmd.OpenQuery "A_C_check_limits_loaded", acNormal, acEdit

' show summary of temp_analyte_determination so we can see if the ranges are wrong

DoCmd.OpenQuery "A_C_check_temp_analyte_determinations", acNormal, acEdit

' append to bgs.dta_analyte_determination database

DoCmd.OpenQuery "A_append temp to AD", acNormal, acEdit

' adds rows to log file.

DoCmd.OpenQuery "A_append temp to AD_log", acNormal, acEdit

' finished

Beep

MsgBox "complete!", vbOKOnly, ""

```
append_analyses1_Exit:
   Exit Function


append_analyses1_Err:
   MsgBox Error$
   Resume append_analyses1_Exit
```

End Function


```
'---------------------------------------------------------
```
' **append_analyses_open_loading_log1**
'
```
'---------------------------------------------------------
```
Function append_analyses_open_loading_log1()
```
On Error GoTo append_analyses_open_loading_log1_Err


   DoCmd.OpenTable "LOADING_LOG", acNormal, acEdit


append_analyses_open_loading_log1_Exit:
   Exit Function


append_analyses_open_loading_log1_Err:
   MsgBox Error$
   Resume append_analyses_open_loading_log1_Exit
```

End Function


```
'---------------------------------------------------------
```
' **append_analyses__2**
'
```
'---------------------------------------------------------
```
Function append_analyses__2()
```
On Error GoTo append_analyses__2_Err


   If (compare_database("a_same_rows_as_database") = True) Then
      ' checks if data already loaded
      Beep
```

MsgBox "if rows are shown then these have the same primary key as in the geochemistry database. (i.e data already loaded) ", vbOKOnly, ""

   End If

append_analyses__2_Exit:

   Exit Function

append_analyses__2_Err:

   MsgBox Error$

   Resume append_analyses__2_Exit

End Function

'--

' **append_analyses__3**

'

'--

Function append_analyses__3()

On Error GoTo append_analyses__3_Err

   Beep

   MsgBox "The program will prompt you if you want to append rows . At the end of the appending new rows process, some rows may be detected as already in the database.", vbOKOnly, ""

append_analyses__3_Exit:

   Exit Function

append_analyses__3_Err:

   MsgBox Error$

   Resume append_analyses__3_Exit

End Function

Appendix 3   LOADER Help information

If detection limit codes are not present, null entries are added to the dta_analytes (or dta_raw_analytes if loading raw data) to satisfy the database rules.
raw data is assumed to be in mg/kg unless the analyte is an oxide, and in this case the data is assigned % as the unit of measurement
Soil horizon_type is by default loaded as 'undifferentiated'.
normal Site description codes(landuses,etc) are loaded for both 'S' and 'A' sites. The 'rows loaded for 'S' sites require an 'A' was collected.(this is field practise anyway)
Contamination codes are assigned to the site it is observed at. Contamination in 'A' soil sites is assumed to be in the 'S' soil site as well.
When you import excel into access , data types are decided automatically based on the first values it finds in the top 24 rows. If none is found, DOUBLE is wrongly chosen. Make sure the fields are
This program loads GBASE field card data and can also load analytical GBASE data:
below is a summary of W:\IM\corporateISmaintenance\Code\Geochemdatabase\field_and_analyses\bulk_loading.doc
ANALYTICAL G-BASE DATA
You have to link from within ACCESS to the data you wish to load using the menu option....FILE--->get external data
Make sure you have the same column names as those required, i.e project,siteno,sample_type,duplicate,method,batch_id,lab plus any number of valid analyte column names.
Column names for analytes must match those listed in the ' view_valid_codes + column names' section of the form (except that occurences of '.' within analyte names must be replaced by '_')
The loading program will check if you have got duplicate rows, invalid values and stop the program if there are.
The program allows data that is in Percent Oxide to be loaded and converted to ppm of analyte. Only Data whose column names indicate it is Oxide data are converted.
Your data is copied to a temporary table whose structure mimics the geochemistry database table, and its contents (if modified data) are compared to the main geochemistry database.
Records of how many rows are transferred to the geochem database are appended to the LOADING_LOG table
Note that the METHOD field value must apply to all the analyte values on the same row. Extra rows must be added if this is not the case.
All data values are loaded. If they have qualifiers, they must precede the value and be one of the valid codes used in mta_dom_analysis_qualifier.
G-BASE FIELD CARD DATA LOADING
This requires your datafile to be W:\IM\corporateISmaintenance\Geochemdatabase\field_and_analyses\field_data.xls and the column names must not be changed.
Columns don't have to be in a specific order. Null entries are entered as nulls.
'S' samples (for modified data only) have 50000 added to sitenos automatically since GBASE doesn't record sites at different depths with different sitenos. If sites at depth ever given same siteno at a
If only the bottom depth of the sample site is given then 15cm is taken off to give the top of the sample.
The project and siteno fields need to be unique for every row. By default, 50000 is added to the 'S' samples. This is also applied to the analyses.
The facility for loading analytical data will ask whether dataset is raw or modified and then load to appropriate tables.
The linked table must be called 'MYDATA'. Overwrite existing table. Data will be appended to DATA_TO_LOAD and must have same column names, and data type , but not all needed.
If loading raw data then the list 'analyte_raw' accessed from the 'view valid codes' button is to be used
detection limits are to be entered on the first row of the data and have a siteno < -1. limits assumed to be the same for all the values in a data column .

References

Most of the references listed below are held in the Library of the British Geological Survey at Keyworth, Nottingham. Copies of the references may be purchased from the Library subject to the current copyright legislation.

AULT, L, MACKENZIE, A C, INGHAM, M N, VICKERS, B P, GRIMSLEY, L D and JOHNSON, C C 2006. From LIMS to Geochemistry Database: G-BASE Samples Analytical Data. *British Geological Survey*, BGS Internal Report Number IR/06/075.

COATS, J S. 2004. The BGS Geochemistry Database: history, design and current usage. *British Geological Survey*, Internal Report Series IR/04/033.

HARRIS, J R, and COATS, J S. 1992. Geochemistry databse: data analysis and proposed design. *British Geological Survey Technical Report*, WF/92/5.

JOHNSON, C C. 2005. 2005 G-BASE Field Procedures Manual. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/05/097.

LISTER, T R, and JOHNSON, C C. 2005. G-BASE data conditioning procedures for stream sediment and soil chemical analyses. *British Geological Survey*, BGS Internal Report Number IR/05/150.

LISTER, T R, FLIGHT, D M A, BROWN, S E, JOHNSON, C C, and MACKENZIE, A C. 2005. The G-BASE field database. *British Geological Survey*, BGS Internal Report Number IR/05/001.

MACKENZIE, A C. 2002. G-BASE bulk data loading application (LOADER). Unpublished Report, British Geological Survey. Last updated 13/09/05.