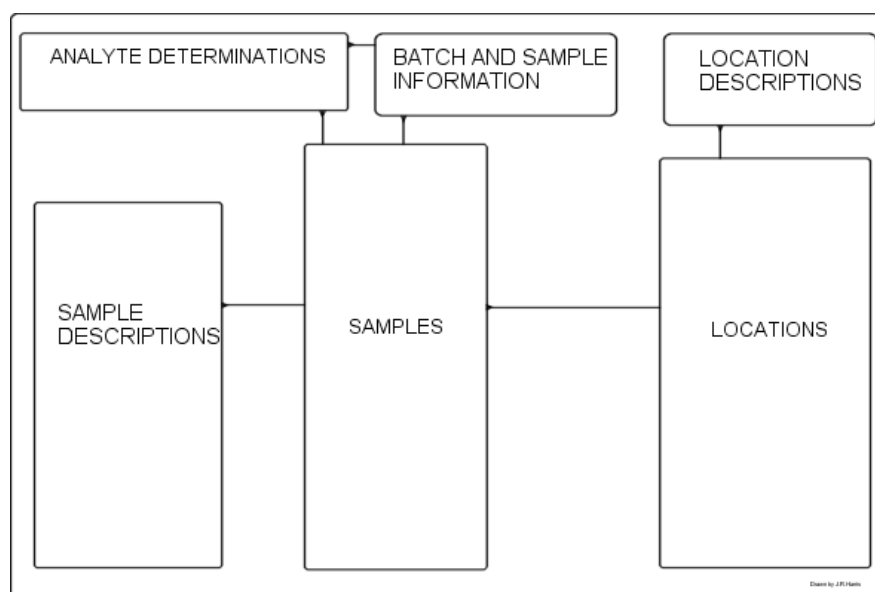




# The BGS Geochemistry Database: history, design and current usage

Economic Minerals and Geochemical Baseline Programme

Internal Report IR/04/033





BRITISH GEOLOGICAL SURVEY

INTERNAL REPORT IR/04/033

# The BGS Geochemistry Database: history, design and current usage

J S Coats

The National Grid and other  
Ordnance Survey data are used  
with the permission of the  
Controller of Her Majesty's  
Stationery Office.  
Ordnance Survey licence number  
GD 272191/1999

## *Key words*

Geochemistry; database;  
information technology.

## *Front cover*

Schematic entity – relationship  
diagram for the Geochemistry  
Database

## *Bibliographical reference*

COATS, J S. 2004. The BGS  
Geochemistry Database: history,  
design and current usage. *British  
Geological Survey Internal  
Report*, IR/04/033. 29pp.

## BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham and Edinburgh; see contact details below or shop online at [www.thebgs.co.uk](http://www.thebgs.co.uk)

The London Information Office 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.*

### **Keyworth, Nottingham NG12 5GG**

☎ 0115-936 3241 Fax 0115-936 3488  
e-mail: [sales@bgs.ac.uk](mailto:sales@bgs.ac.uk)  
[www.bgs.ac.uk](http://www.bgs.ac.uk)  
Shop online at: [www.thebgs.co.uk](http://www.thebgs.co.uk)

### **Murchison House, West Mains Road, Edinburgh EH9 3LA**

☎ 0131-667 1000 Fax 0131-668 2683  
e-mail: [scotsales@bgs.ac.uk](mailto: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: [bgs london@bgs.ac.uk](mailto:bgs london@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, 20 College Gardens, Belfast BT9 6BS**

☎ 028-9066 6595 Fax 028-9066 2835

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

☎ 01491-838800 Fax 01491-692345

### *Parent Body*

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

☎ 01793-411500 Fax 01793-411501  
[www.nerc.ac.uk](http://www.nerc.ac.uk)

# Acknowledgements

A large number of individuals in BGS have contributed to the Geochemistry Database and this assistance has continued over the long period of the database's operation. In addition to those given acknowledgement in the references, special mention should be made of the contribution of John Harris who first began the redesign of the Geochemistry database in 1990. Further acknowledgements should be given to Claire Nicholson, Gaynor Collins and Sue Hobbs who have kept the database running smoothly over the past years. Alan Mackenzie has also been one of the key individuals working on the ORACLE database and his contribution has been, and still is, invaluable. Several Sandwich students worked on the database during the 1990s and their contribution is acknowledged in the list of database reports they helped to produce.

## Contents

<b>Acknowledgements.....</b>	<b>i</b>
<b>Contents.....</b>	<b>i</b>
<b>Summary .....</b>	<b>ii</b>
<b>1 Introduction .....</b>	<b>3</b>
<b>2 Analytical Chemistry .....</b>	<b>5</b>
2.1 Detection limits.....	6
2.2 Methods employed in G-BASE and MRP .....	7
2.3 Reporting .....	9
2.4 LIMS.....	10
<b>3 Database Design .....</b>	<b>10</b>
3.1 Early beginnings .....	10
3.2 G-Exec Database System.....	11
3.3 Oracle.....	12
3.4 The 1992 Design.....	12
3.5 Current Structure .....	15
<b>4 Current Usage.....</b>	<b>16</b>
4.1 Frequency .....	16
4.2 Data Retrieval .....	16
<b>5 Problems.....</b>	<b>19</b>
5.1 Data problems .....	19
5.2 Design problems .....	19
5.3 Staff problems.....	19
<b>6 Conclusions and Recommendations .....</b>	<b>20</b>
6.1 Future Developments and Recommendations .....	21

<b>Glossary.....</b>	<b>23</b>
<b>References .....</b>	<b>24</b>

## FIGURES

Figure 1 Comparison of the numbers of laboratory numbers analysed by AAS and XRF.....	7
Figure 2 1970 field card design for drainage samples .....	11
Figure 3 Schematic Entity – Relationship diagram for the Geochemistry database.....	13
Figure 4 Detailed Entity – Relationship diagram (after Coats and Harris, 1995).....	14
Figure 5 Detailed Entity – Relationship diagram for the current implementation of the Geochemistry Database.....	16
Figure 6 Example of geochemical data on the GDI from Tiree, Scotland showing the location of MRP rock and G-BASE stream sediment samples. ....	18
Figure 7 Example of geochemical data available internally via the GDI.apr GIS application .....	18

## TABLES

Table 1 Analytical methods employed in the BGS Laboratories.....	5
Table 2 Number of laboratory numbers completed in year by each analytical method.....	8
Table 3 Methods of analysis and number of element determinations ranked in order of importance .....	9

## Summary

This short report provides an introduction to the history of the BGS Geochemistry Database and traces its origins back to 1970 when the first field cards were used by the forerunner of EMGB. Subsequent data collection by the Mineral Reconnaissance and the Geochemical Baseline Programmes led to the creation of large data files in the BGS G-EXEC data system and in simple ORACLE databases. This data was incorporated into a unified Geochemistry Database, the design of which was proposed in 1992.

The present Geochemistry Database largely follows the logical design put forward in 1992 by Harris and Coats but a number of changes are listed which were implemented in ORACLE in 1993. These changes, along with some relatively minor changes carried out subsequently, are discussed and the reasons given for their adoption.

Information Systems have changed markedly over the last ten years and the current Client – Server model used for interrogating the database would have been previously impossible because of the lack of desktop power. Five methods of linking to the database using this model are described along with their capabilities. A number of problems or defects in the database are identified and in the final section some conclusions and recommendations are presented for consideration. A full bibliography of the internal reports and publications on the database is given along with a glossary of technical terms.

# 1 Introduction

This introduction aims to provide a brief history of the use of geochemistry in BGS and the political and organisational framework, which led to its development and use. The author was involved in some of the work, initially in the analytical laboratories and then in the Mineral Reconnaissance Programme and this history is inevitably a subjective one.

BGS and its forerunners have used traditional whole rock geochemistry since the 19<sup>th</sup> Century to classify igneous rocks, but the use of drainage geochemistry in the search for unknown and unexposed mineral deposits dates back firstly to the work of the Overseas Geological Survey in the 1950s and 1960s. Mining companies had used soil geochemistry in countries, such as Zambia, to locate new ore deposits in the 1950s (Hawkes and Webb, 1962) and Imperial College had a very successful research team working in this evolving discipline. On a personal note it should be mentioned that classical or ‘pure’ geochemists looked down somewhat on the activities of ‘applied’ geochemists who analysed weathered or superficial material and it was considered a less scientific activity than the analysis of pure and unaltered silicate rocks. This attitude still survives to some extent in BGS. The Overseas Geological Surveys (OGS) had their own laboratories, firstly based at the Imperial Institute in South Kensington but then relocated to Grays Inn Road in 1960s. The OGS was merged with the Geological Survey of Great Britain to form the Institute of Geological Sciences (IGS) in 1966. Thus BGS, or its forerunner IGS, gained the experience in the analysis of surficial materials for the purposes of mineral exploration.

The first programme to use drainage geochemistry in the UK was the Uranium Programme carried out by BGS for the United Kingdom Atomic Energy Authority (UKAEA) between 1966 and 1972. Earlier surveys for uranium and rare minerals (hence an earlier name for the unit carrying out this work, the Radioactive and Rare Minerals Unit RMMU) had used geochemistry but only to determine the grade of rock or ore samples, not for systematic exploration. It did not at that time routinely analyse surficial materials, such as soils, tills or drainage samples. The new Uranium Programme was the first BGS project to analyse waters, soils and drainage sediments for mineral exploration. Systematic surveys for uranium mineralisation were carried out over large areas of Northern Scotland, most notably over the Orcadian basin of the Orkney Islands, Caithness and Sutherland. Other surveys were carried out in the English Midlands and South West England, but in the latter area only localised surveys were conducted around the known uranium mineralisation. Regional surveys grew out of these uranium surveys and the first regional geochemical drainage sampling projects took place in South Orkney and Caithness in 1969 (IGS, 1979), followed by Shetland (IGS, 1978a) and the rest of Orkney in 1970 (IGS, 1978b).

The mid- 1960s was a period of notable advances in analytical chemistry. Until that time most trace element analysis was done by labour intensive colorimetric methods or by optical emission spectrometry (OES), but in the period 1960-1970 a number of new techniques, such as atomic absorption spectrometry (AAS), direct-reading emission spectrometry (DRES) and X-ray fluorescence spectrometry (XRF), became commercially available. These allowed the more rapid and automated determination of a wider range of elements in geological materials. The older methods were, however, not completely superseded until the late-1970s. This period of rapid technological advancement (called ‘the white-hot heat of technological revolution’ by the government of the time) coincided with a period of great expansion in IGS’s role and size. It grew rapidly in staff numbers and moved into several new geoscientific disciplines. The analytical laboratories expanded in size and purchased a wide range of new equipment.

Several people were instrumental in guiding the expansion of geochemistry over this period: Sir Kingsley Dunham, the Director of IGS, who was an economic geologist and saw that these methods opened up new ways of locating mineral deposits in the UK; Stan Bowie, who was head

of the newly formed Geochemistry Division, which amalgamated the OGS analytical and mineralogical laboratories with the RRMU, and Denis Ostle who led the latter unit. Jane Plant was one of the graduates recruited at this time and she became increasingly involved in the geochemical surveys of Northern Scotland.

The Uranium Programme was designed to locate indigenous supplies of uranium to the growing number of UK nuclear reactors and to the nuclear weapon stockpile but by 1974 new discoveries of uranium in Canada, Namibia and Australia had reduced the need and the programme was terminated. Stan Bowie faced with a severe reduction in funding lobbied the government and the newly formed Department of Industry agreed to fund two new programmes, the Mineral Reconnaissance Programme (MRP) and the Regional Geochemistry Research Programme (RGRP). The two programmes were both in the same unit, re-christened from RRMU to firstly the Radioactive and Metallic Minerals Unit (RMMU) and then to the Metalliferous Minerals and Applied Geochemistry Unit (MMAGU). There was some inter-changeability between the two programmes and the geochemical methods were essentially the same. These were largely based on Plant's 1971 orientation work (Plant, 1971), which studied geochemical dispersion downstream of known deposits, and work by Gallagher, Michie, Cooper and others on the Uranium programme (Gallagher et al, 1971). The MRP and RGRP were complementary in that the MRP could follow-up anomalies discovered by the RGRP and any MRP regional surveys could be incorporated in the regional geochemistry.

Because the RGRP was a regional survey, which was envisaged to cover all of the UK, its field sampling and recording were more systematic than the more ad-hoc problem-orientated methods employed by the MRP. Sets of drainage field cards were developed by staff, initially based on port-a-punch cards but then standardising on filofax cards, which were later keyed on to IBM 80-column punched cards. As the sampling was carried out by largely untrained students data entry was strongly formalised with a fixed range of entries allowed. These field cards formed the basis for the geochemistry database. Similar cards were developed by the MRP for other sample types such as soil, rock and drill core. These developments in field recording were matched in the analytical laboratories where the first computer an IBM 1130 was installed by the author in 1969. Analytical determinations were initially recorded on Fortran forms and punched on to cards, but over the next five years direct data capture by the computer from the Direct Reader analytical instrument was implemented.

As the RGRP progressed south through Scotland it began to overtake and incorporate MRP surveys but as time went on the two diverged in several key areas of methodology. Panned concentrates became more important to the MRP than the standard stream sediment sample (Leake and Smith, 1975) and there was less need for the MRP to do regional sampling. The need for deeper till sampling in the glaciated Highlands also focussed the MRP on smaller areas, which culminated in the discovery of the Aberfeldy barite deposits in 1979. The divergence was also accelerated by personnel changes and the move of the MRP staff to Keyworth in 1978-80. This divergence led to slight differences in field techniques and different analytical methods being employed, with the MRP increasingly using XRF and the RGRP relying on the direct reading emission spectrometer (DCOES) methods. This increased divergence led to several other changes in management and the RGRP became its own Group and even moved into a different BGS Division to the MRP. With the formation of the Economic Minerals and Geochemical Baseline Programme (EMGB) in 2001 the two projects, now called the Minerals and the G-BASE Programmes, become reunited.

Geochemistry continued to be used by other groups of staff in BGS. The geological mapping staff continued to send a small number of samples for major element analysis. Numbers are difficult to estimate but they probably submitted in the order of 100 samples for analysis per year, whereas the MRP and RGRP collected several thousand samples per year. Some major regional mapping projects in the 1980's, such as the Lake District and Snowdonia projects collected a larger number of rock samples and geochemists trained in the MRP were largely



employed to collect samples and to interpret these results. Other BGS groups, such as the Industrial Minerals Assessment Unit, also collected geochemical data on limestones as part of surveys of the Peak District and Mendips for high purity limestone resources.

Internationally there were a number of mineral exploration and mapping projects carried out by BGS in the 1960-1980s. Much of this work was funded by the Overseas Development Agency and, later, by the World Bank. The analysis was initially carried out in the IGS laboratories but increasingly in laboratories set up in the host countries. Some of these surveys were very large, collecting many thousands of drainage samples. The geochemical data has been retained by BGS in many cases and is available as stand alone files but is not incorporated into any larger unified database.

## 2 Analytical Chemistry

As discussed earlier analytical methods used by BGS have changed and improved over the 35 years from 1968 when systematic sample collection began. Table 1, which is a copy of the database table BGS.MTA\_DOM\_ANALYSIS\_METHOD, lists the methods which have been employed by BGS. The majority of the analyses in the Geochemistry Database have been performed by the BGS analytical laboratories but outside laboratories have been used, particularly for gold and the platinum group elements where the analytical methods were not available internally. The laboratories are listed in the database table BGS.MTA\_DOM\_ORGANISATION.

Table 1 Analytical methods employed in the analysis of BGS samples

Code	Translation	Defined as
AAS	Atomic Absorption Spectrometry	Atomic Absorption Spectrometry.
AFS	Atomic Fluorescence Spectrometry	Atomic Fluorescence Spectrometry
AQTOC	Shimadzu 5000	
BERYL	Beryllometry	Gamma excitation of Beryllium and detection of neutrons.
BETAP	Beta Probe	Direct electron excitation X-ray spectrometry.
CE1500	CARLO ERBA 1500 CHNS Analyser	
COL	Colorimetric Method	Colorimetric Method
COLHEX	Colorimetric Hot Extractable	Colorimetric Hot Extractable
CVAAS	Cold Vapour AAS	Cold Vapour Atomic Absorption Spectrometry
DCOES	Direct Reading OES	Direct Reading Optical Emission Spectrometry using photocells to record the light intensity at different wavelengths
DNA	Delayed Neutron Activation	Delayed Neutron Activation
ETAAS	Electrothermal AAS	Electrothermal Atomic Absorption Spectrometry
FAAS	Fire Assay AAS	Fire Assay Atomic Absorption Spectrometry
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
ICPAES	ICP Atomic Emission Spectrometry	Inductively Coupled Plasma Atomic Emission Spectrometry
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
LECO	LECO CHNS Analyser	LECO CHNS Analyser
LIFS	Laser Induced Fluorescence Spectrometry	Laser Induced Fluorescence Spectroscopy using Scintrex machine
LOI105	Loss on ignition at 1050°C	Gravimetric method to determine organic carbon content
LOI450	Loss on ignition at 450°C	Gravimetric method to determine H2O content
NA	Neutron Activation	Neutron Activation
NPOC	Organic Content	Measurement by non-dispersive infra red of the CO2 from combusted sample that had been acidified and put on Pt Column at 680c
OES	Optical Emission Spectroscopy	Optical Emission Spectroscopy using photographic plates to record the spectra
PE240	PERKIN-ELMER 240 CHNS Analyser	
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

The analysis method Code specifies the analytical method used for every determination and is stored in each record of the Analyte Determinations table. This is especially important where the analytical method is a partial or selective in its extraction of the element. It is also possible to hold in the database the analytical determination by two or more methods on the same sample. In providing both a date of registration and a method it is possible to ascertain the complete procedure used in the analysis if the laboratory retains these records. Also different analytical methods have different strengths and weaknesses so that Co by XRF is less reliable than Ni because of the interference by Fe, whereas this does not hold true for AAS.

## 2.1 DETECTION LIMITS

Detection limits are stored in the Geochemistry Database in table BGS.DTA\_ANALYTES. However, this information was rarely entered because much of the data, particularly that

collected by the MRP, was used as a *comparative* measure of element abundance, which for many surficial samples was adequate for the task. *Absolute* values of element abundance and comparison with international standards were considered of lesser importance. The G-BASE programme needs consistent and reproducible element values across the country but this is difficult to achieve where different analytical methods have been used. For example, the early AAS methods only gave a partial extraction of the trace elements Cu, Pb and Zn, whereas the later XRF method provides an analysis of the total metal content. The table with detection limits is linked to the BGS.DTA\_BATCHES and the BGS.DTA\_ANALYSIS tables because every sample should have a batch number, which identifies the reporting date of the analysis and an analytical method.

An *ad hoc* determination of the detection limit can be estimated by plotting log normal – cumulative frequency graphs because the analytical variation normally has a normal distribution whereas the geological variation normally shows a lognormal one. However this only provides a broad estimation of the detection limit which is better obtained by repeated analysis of a known standard.

## 2.2 METHODS EMPLOYED IN G-BASE AND MRP

The analytical methods and detection limits are given in the Geochemical Atlases for the G-BASE programme and summarised by Johnson and others (2004). Methods used MRP projects are described in the MRP reports (1-146) and Data Releases (1-23) but they do not always give details of detection limits for reasons discussed above..

The number of determinations is illustrated by Figure 1, which shows the number of batches (Laboratory Numbers) completed by AAS and XRF in each year for the period 1970-2000.

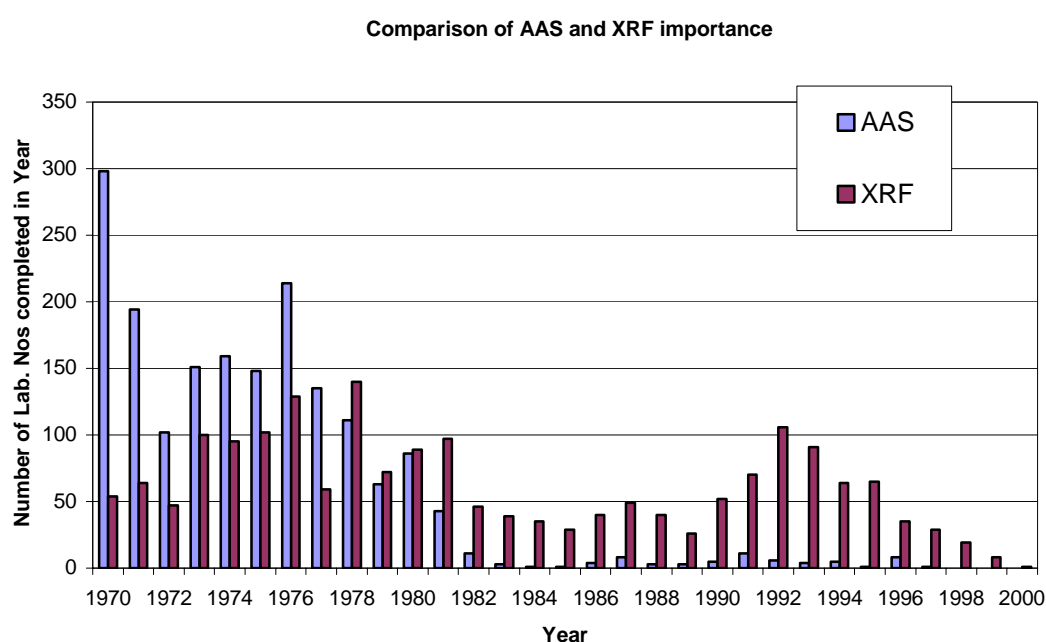


Figure 1 Comparison of the numbers of laboratory numbers analysed by AAS and XRF

The methods used can be summarised as follows:

Method	Typical Elements	Period used
AAS	Cu, Pb, Zn $\pm$ Co, Ni, Fe, Mn	1968-1982
OES (Spectrograph)	B, Ba, Be, Co, Cr, Fe, Mn, Mo, Ni, Sn, Ti, V and Zr	<1968-1975
DCOES (Direct Reader)	Multi element	1974-1992
DNA	U	<1968-1992
XRF	Multi element	1970-2003

Table 2 Number of laboratory numbers completed in year by each analytical method

METHOD 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000

AAS	298	194	102	151	159	148	214	135	111	63	86	43	11	3	1	1	4	8	3	3	5	11	6	4	5	1	8	1			
AQTOC																											3				
BERYL				1																											
BETAP				2	2	2	12	83	42	33	25	65	30	18	9	2															
CE1500																							1								
COL	71	62	46	59	59	31	87	46	22	16	45	20	1	2	2	1	1														
COLHEX											3																				
CVAAS																												1			
DCOES	17	10	12	14	21	8	9	10	6	4	12	10	2	1	2	1	1	1	2	1	2	2									
DNA	355	241	110	124	103	58	30	29	49	17	19	2	7	2	1	2	1	1	1	1											
FAAS															2	1	1	15	7	10	1	2									
FDCP																															
FICP																				6	1	1	4								
FICPMS																		1	13	3	1										
FLAES				1		3																									
GRAV				3	9	1	11	77	39	24	15	51	23	1	3	1	1	8	5	7	4	1									
HYICP																				5	5										
ICP											1									7	1	2	2	6	1	2	5				
ICPAES																							5	15	3	5	3				
ICPMS																								4	3	4	3	1	1		
ISE								2	2	3	2	1	1	1	1	1															
LECO																															
LIFS				3													1			1											
LOI450																									1			1			
NA																							1								
NPOC																										1	1				
OES	68	86	75	93	65	37	30	23	38	42	28	12	3	1	1																
SPPHOT					2	3																									
TITRE																								1	2						
XRF	54	64	47	100	95	102	129	59	140	72	89	97	46	39	35	29	40	49	40	26	52	70	106	91	64	65	35	29	19	8	1
XRFB											1		2	9	6	8	4	2	1	6	17	62	47	16	12	10	6	7	4		

In terms of the numbers of samples analysed Table 3 shows that direct current optical emission spectrometry by direct reader (DCOES) has been used to produce about 2.3 million element determinations and XRF slightly less at 2.2m. These two methods account for around 80% of the total number of analyses.

Table 3 Methods of analysis and number of element determinations ranked in order of importance

METHOD	Number of element determinations
DCOES	2295865
XRF	2193829
AAS	313207
OES	256377
ICPAES	225091
DNA	135096
ICPMS	52968
ICP	43278
XRFB	39133
ISE	28796
COL	28605
BETAP	23281
FAAS	10895
FICPMS	8532
GRAV	7893
LIFS	6721
FICP	1483
HYICP	1440
NA	1344
NPOC	1260
LOI450	1066
COLHEX	596
<b>Total</b>	<b>5676756</b>

Only one Laboratory Number was registered in 2000 and no subsequent entries have been made to the table BGS.DTA\_BATCHES, which holds the Laboratory and Laboratory numbers the date registered and date reported. This table is important for metadata and to enable the original sources of the results to be located. It may be the case that the tables have been superseded by the LIMS but in which case there should be equivalent, easily available metadata from LIMS.

## 2.3 REPORTING

Data has been reported from the analytical laboratories in a variety of forms, ranging from hand written columns of data on paper to electronic files on diskette or attached to e-mails. In the 1960's most data was reported in hand written or typed reports and a copy was kept in the filing system at GIR. The geologist requesting the analyses if they wanted to use a computer for interpretation or plotting rewrote them on to Fortran forms. These were A4 sized sheets with an

80-column grid so that data could be keyed onto 80-column punched cards. Other variants of punched cards, such as port-a-punch and mark sense cards, were tried but they both proved unsatisfactory. This process carried on until the advent of computers or microprocessors attached to the analytical instrument. Some of these produced punched tape, which could be read directly into the computer or, as in the case of the Direct Reader DCOES, the instrument was directly connected to the computer over a custom built serial link. The first BGS XRF instrument produced data typed on continuous roll feed paper, which had to be re-entered on to Fortran forms. Results were normally reported in batches of each Laboratory number and method, so that if AAS, DNA and DCOES were requested these would be reported on three separate sets of forms, which may arrive at various times depending on the backlog in the laboratory. The requesting geologist carried out integration of the data.

With increasing computing resources in the laboratory results were transferred on paper tape, punched card and magnetic diskettes, originally in 8", then 5.25" and 3.5" sizes. These diskettes and the punched cards formed the original source documents along with the paper copies and examples of these exist around the present BGS offices. Punched card and 8" diskettes are obsolete and cannot be read by PCs; and 5.25" diskettes can only be read by a few computers. Later when network facilities became available data transfer was completely electronic to a central computer from which the file was transferred to the geologist.

The original source data of the analytical results reside in a variety of formats; exclusively on paper, either hand written on paper or forms, as machine produced typed data and in a variety of electronic formats. To prevent data loss or erosion it is important to keep these original records in a secure archive, a secure database, or, preferably, both.

## **2.4 LIMS**

Many of the metadata functions of the database are now incorporated in the LIMS and it was realised in 1992 that the database if it was to manage all the data recording and management functions of the MRP and G-BASE data and the laboratory it would become a LIMS. This was clearly beyond the scope of the database and so only a limited functionality was built in to the Geochemistry Database and that mainly for historic data which would never be subsequently added to the LIMS. However since the LIMS has been operational there has been a breakdown in the flow of some critical information into the Geochemistry Database and it is important that this flow becomes re-established as it provides the only metadata accessible at the same time as the data.

# **3 Database Design**

## **3.1 EARLY BEGINNINGS**

The first part of a database design for geochemistry was the design of field cards for the collection of field data on stream, soil and rock samples. The first set of these was designed in 1970 and these cards, along with the detailed instructions were issued to all field sampling teams. These cards because they were designed to hold all the field information collected by the MRP and regional geochemical teams were an effective and efficient database, structuring the field data collected in a systematic way. An example of a drainage card from 1970 is shown in Figure 2. At the same time a standard numbering system for geochemical samples was devised, which was clearly essential with the increased numbers of samples collected and analysed. A detailed description of all the field cards from 1970 to 1992 is given in Harris and others (1993), which also provides a translation to the 1992 database codes. Copies of all the blank field cards and related instructions for use have been filed by Sue Hobbs.



BGS decided that it needed to centralise their expertise and a small Computer Unit was formed which consisted originally of just three people, Vic Loudon, Clive Jeffery and Dave Farmer, assisted by some data preparation staff. Many of the tasks performed by BGS computer users were fairly standard and the Computer Unit wrote a suite of application programs for these common tasks. This application suite grew into a combined application and database package called G-EXEC.

G-EXEC was a file-based package in which the data files contained a header section which self-described the contents of the file, so that numeric and text fields were correctly identified. It was possible to do simple query functions, and link, merge and split files etc. It was called a database system and it bears some similarities to dBase in its table structure but, essentially, it was an application package with some added database functions.

G-EXEC could not handle null values but represented these by -9999.0 or similar codes and even though all the data entry was on 80-column cards the files often contained in excess of 30 columns of data values, once the analytical was combined with the locational data. Large G-EXEC datasets were prepared for the Geochemical Atlas areas of Shetland, Orkney, Southern Orkney and Caithness, Hebrides and Great Glen. Similar compilations were made in MRP for large areas such as South West England, Exmoor, North Wales, Anglesey and parts of the Dalradian. Most of these datasets only incorporated the analytical and a limited part of the field data. G-EXEC was very good at handling numeric data but less efficient and powerful at manipulating text data.

G-EXEC was used in BGS from about 1973 until 1985 and a cut-down version was installed on the GEC 4090 and VAX 8600 computers installed in Keyworth. BGS purchased the ORACLE database software around 1985 and G-EXEC fell out of use. Most of the geochemical data held on magnetic tape in G-EXEC text format was transferred to ORACLE database software in 1986 onwards.

### **3.3 ORACLE**

ORACLE database software was purchased by BGS in the mid-eighties and Keith Holmes and the author designed database tables that essentially replicated the structure of the G-EXEC data files. This un-normalised MRP database was loaded with the large datasets such as South West England and the others mentioned in the previous section. John Durham designed a similar but separate database for the G-BASE data. During 1988 a menu-driven front end was developed that allowed non-expert users to access the data and a rock sample table was established.

These databases carried on until the early 1990s during which time the Argyll and East Grampians Atlases were published. However the structure of the databases was becoming very difficult to manage and some data, such as that on soils, was not incorporated. By the summer of 1989 it was obvious that a new data analysis was required to rectify these defects and to take account of data types not identified in the original database design. This data analysis was carried out between 1989 and 1991 and a detailed description of all the MRP data fields was created. These produced initially a set of 1990 data codes into which all the previous field records could be amalgamated. During this process it was realised that the G-BASE data was essentially compatible with the MRP because of their common origin. Whilst the regional geochemical programme had evolved away from the MRP they could be compared to sub-species rather than full species and were fully capable of integration or interbreeding. Merging the two ORACLE databases had several advantages and the proposal to create a single BGS Geochemistry database was accepted by BGS management in 1992. A full report on the logical data design was published by Harris and Coats in 1992 and a further ten reports published in the Applied Geochemistry Series.

### **3.4 THE 1992 DESIGN**



The logical design of the BGS Geochemistry Database was published by Harris and Coats in 1992 as MRP Report No.125 and the reader is advised to read that report which explains in detail the design that as subsequently implemented in the following two years. The following text is a summary of that report covering the essential features that current database users require. It should be noted at this point that a large volume of geochemical data had been collected over the period 1965 to 1992 and essentially the database design was data driven. The design was therefore not highly theoretical or completely normalised to the highest degree but a practical one, which would accommodate the maximum amount of the existing data at the least expense. Some of the later problems with the design stem from this compromise between practicality and database theory.

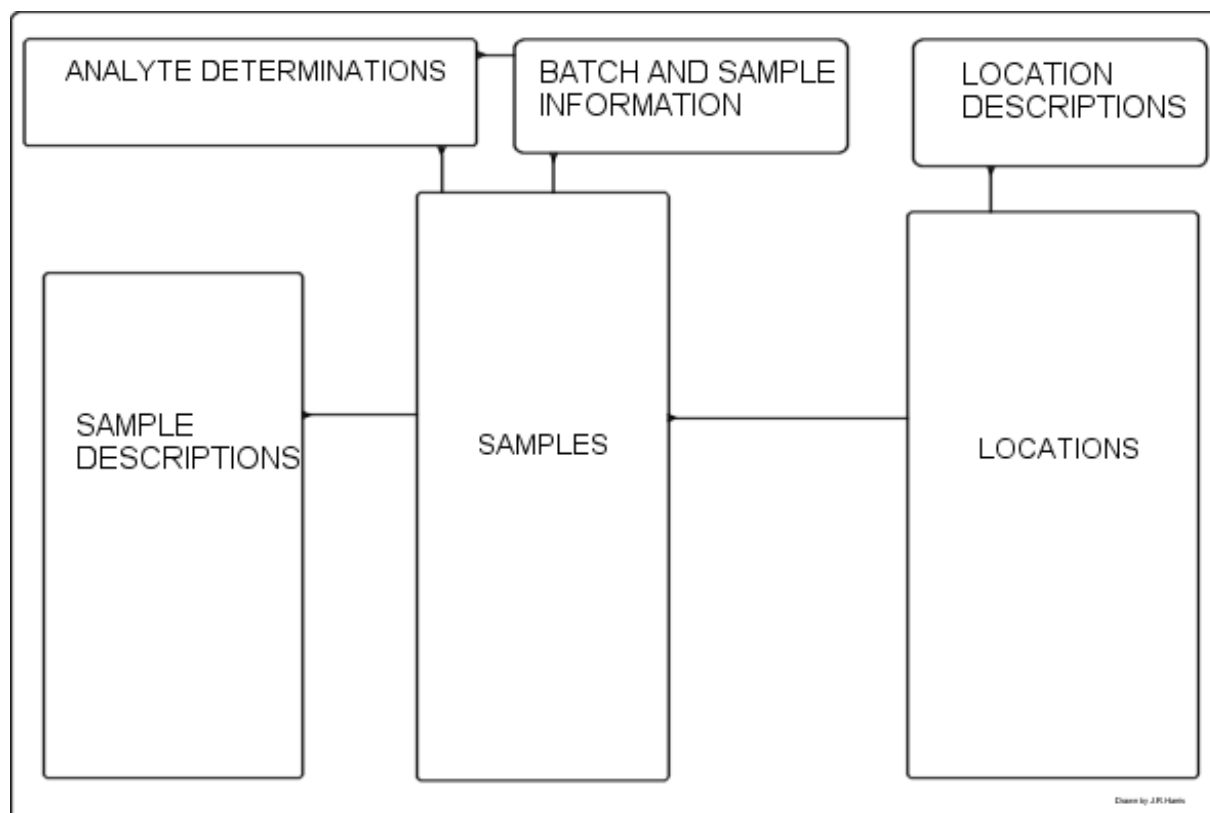


Figure 3 Schematic Entity – Relationship diagram for the Geochemistry database

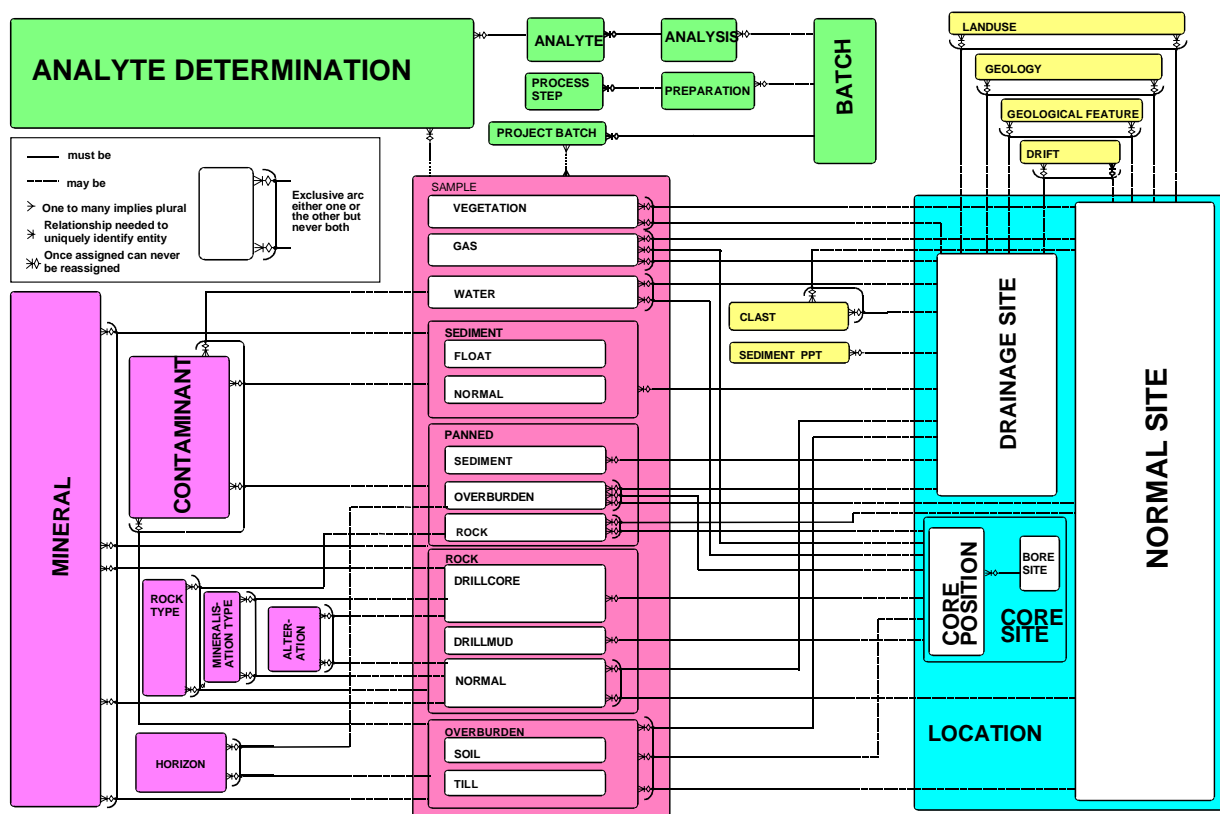


Figure 4 Detailed Entity – Relationship diagram (after Coats and Harris, 1995)

The schematic relationship between the database entities is shown in Figure 3 and a more detailed one, which can be mapped to the physical model in Figure 4.

The underlying principle of the design is that information is collected on the Location, the Samples and the Analyses. The Location tables describe the position of the site, elevation, top and bottom depths below surface, grid accuracy and map used, collector and date collected and comments. The individual sample tables, Vegetations, Waters, Sediments, Panneds, Rocks, Drillcore, Drillmuds, Normal rocks and Overburdens contain information about the sample such as the colour, mesh size, volume, etc. Because some of these attributes are not applicable to all sample types the fields of each sample table vary. For example, the attribute mesh size is not applicable to a rock sample and the attribute soil type would not be applicable to a water sample. This explains the somewhat confusing multiplicity of tables and that in the E-R diagram entities, such as Minerals, sometimes connect to master entities such as Overburden and sometimes to sub-entities such as Normal rocks.

Another important principle that was inherent in the design of the original geochemical field cards was that the site location was given a unique number such as CZ 1234 (or 22 1234) and this number was inherited by the sample collected from the site. Sites were located in three dimensions - Easting, Northing and Depth below surface or down the borehole. As each site was unique it was logically not possible to sample exactly the same site twice, so that a second sample had to be given a different site number. Exceptions to this rule were made for drainage samples when the stream sediment and the panned concentrate were collected at the same time and the concentrate panned from a portion of the same sample. The stream water was also deemed to be collected from the same site. Thus C, P and W samples were given the same site number and thus the sample numbers could be CZC 1234, CZP 1234 and CXW 1234. Thus the sample number is a compound key made up of the project code, the site number within that code and the sample type. Unfortunately, this principle was not universally adhered to in the field and

sample numbers CZS 1234 and CZR 1234 might have been collected from different locations. These were defects in the system that were not detected by the database design.

### 3.5 CURRENT STRUCTURE

In the past ten years since the 1992 database model there have been a small number of changes to the database structure. These changes had to be effected when the Geochemistry Database was incorporated into the BGS schema and the data modelled into a CASE tool. The most important changes are listed below.

1. A new table BGS.SITE\_LIST has been introduced to hold a complete list of all geochemical sites – Core site, Drainage site and Normal site
2. A new table BGS.SAMPLE\_LIST has been introduced between the BGS.SITE\_LIST and the individual sample tables, and provides a complete list of all the geochemical samples in the database.

This change in effect removes the master entities shown in pink and light blue in Figure 4. In the initial 1992 model it was logically possible to have a drainage site and a normal site with the same site number (Note that each location site is given a unique number that is inherited by the geochemical sample collected at the site). An external SQL program would be necessary to enforce the rule that each and every site should have a unique number. With the addition of the new entity Site\_list this uniqueness is enforced by the database.

Similarly, it was possible in the 1992 model for the same sample number to exist in two of the sample tables, for example in the Overburden and the Rock tables. Creating a new master entity called Sample\_list prevent the same sample number appearing in the individual sample type tables. Note, also, that the original design had separate entities for Soil and Till samples but in the physical design these two were amalgamated into a single Overburden table. These changes would not have been necessary if the ORACLE software included support for Master and sub-entities. Where the same site number has been used for different samples a set of rules have been devised to aid consistent renumbering of sites and samples. These rules are:

Add 10000 to Soil sites  
20000 to Till sites  
30000 to Boreholes/Drill Core sites  
40000 to Rock sites  
50000 to Drainage sites

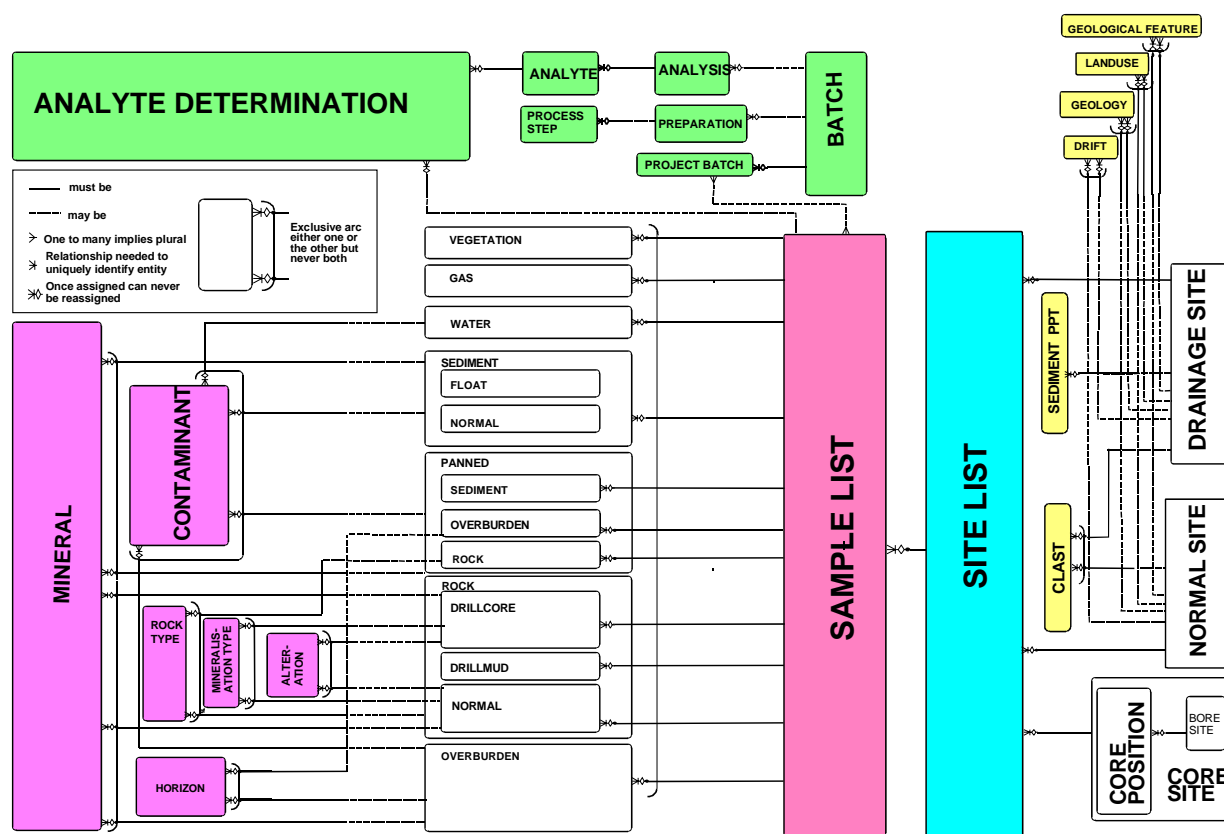


Figure 5 Detailed Entity – Relationship diagram for the current implementation of the Geochemistry Database

## 4 Current Usage

### 4.1 FREQUENCY

A requirements analysis was completed in 1992 (Geochemistry Database report 4, Harris et al., 1992) and this estimated the annual usage of the database. The frequency of use has dropped because of the reduction in staff working on the MRP and its successor the Minerals Programme, and on G-BASE. Also database management has been removed from the former Groups and Division and placed in a separate Information Management Programme along with its funding. Erosion of this funding has taken place by competition for scarce resources and data management of both G-BASE and MP data has been reduced. Also, CR funding from the DTI has been reduced over the years and will cease in 2004 eliminating that revenue source.

Geochemical data sales and licensing have been relatively buoyant due to the increasing importance given to environmental pollution and to the sampling of urban areas. MRP data licences have been relatively few because of cutbacks in commercial mineral exploration, but have recently revived because of the increase in the price of gold. However, all licence revenue goes into a central fund so does not directly benefit the management of the database.

### 4.2 DATA RETRIEVAL

Data in large ORACLE databases can be retrieved in a number of ways. The most direct is SQL\*PLUS where SQL commands are typed into a command line interface and data retrieved to

the screen or to a file. However, there are a number of alternatives, which are simpler and more intuitive to the average user.

#### **4.2.1 SQL\*FORMS**

A data retrieval system was written by Alan Mackenzie (Mackenzie, 1997), which used an ORACLE SQL\*FORMS interface to write SQL commands and retrieve data. The system used Pro\*Fortran to dynamically construct the SQL commands which were sent to the database. Retrievals of analytical results on Atlas area, location codes and sample codes, listed as requirements in Harris et al. (1992) were implemented. The program ran on the VAX computers and whilst reasonably friendly in data input was unable to produce neatly formatted reports. Also the program had to be partly rewritten and recompiled if any of the component software packages changed and it thus had a high maintenance overhead. With the change of computer architecture to a client-server model and more powerful PCs this fell out of use in about 2000.

#### **4.2.2 Microsoft Access**

Microsoft Access is a Windows-based relational database package which is part of the Microsoft Office package installed on the majority of user PCs. It has an interface that is familiar to most users and is a relatively easy product to learn. It readily conforms to the client-server model where the PC acts as a database client and retrieves data from the back-end ORACLE database server. The Open Database Connectivity (ODBC) interface is in fact implemented in a number of database and Office packages. It is inherently more flexible than the compiled database retrieval program such as the SQL\*FORMS based package described above. However, this flexibility in being able to write retrievals using the graphical query interface does imply a greater knowledge of the database structure. This is the preferred interface of the author.

An alternative Access interface has been written by Alan Mackenzie which offers a form based query interface with retrieved data displayed in on-screen forms. This is similar to the SQL\*FORMS design and is used by G-BASE staff to some extent.

Whilst Access is one of the most flexible retrieval systems it requires a certain level of skill and knowledge which beginners find initially daunting. Because the analyte determinations table is fully normalised one, retrieval of multielement data in a flat file, familiar to users used to looking a large spreadsheet, is not immediately obvious and can take some time over a networked PC. Retrieval of a large spreadsheet table with say 15,000 samples for 30 elements, can take several minutes. An un-normalised data warehouse would be the simplest solution but it would be difficult to justify the expense of setting it up and maintaining it with the current level of use.

#### **4.2.3 ArcView**

Other programs can use the client-server model and ODBC to extract data from the ORACLE database and the standard BGS GIS package has this facility. Draft instructions on connecting ArcView to ORACLE databases have been produced by Sue Hobbs. There are restrictions on using this facility, the major one being that you cannot link to ORACLE tables through an Access front-end, you must link directly to ORACLE. It should also be noted that the interface uses SQL queries to link tables and this needs some database experience.

#### **4.2.4 Intranet Data Access**

The IDA (Intranet Data Access) system provides routine searching (browse) and data management functionality for a wide range of BGS data held in ORACLE tables. At present this does not include geochemical data but there is no reason why it could not be extended to data held in the BGS Geochemistry database. The IDA interface is a browser application written in Coldfusion, which takes user input and send an SQL query to the database. Data is returned in a



## 5 Problems

A database is a structured repository for information about the real world and, as it is only a model, it cannot encompass all the different data and the complex inter-relations present in that world. Some data will be found that is too difficult to fit into the model. For example, orientation data where different size fractions of a stream sediment are separated and analysed is very difficult to fit into the Geochemistry Database. The data is usually quite small and it is not usually interpreted along with the rest of the data so it can be safely excluded from the database. So there will always be a few subsets of the data that have to be excluded from the database and these could be described as database problems. Problems can also be caused by incorrect data being entered into the database.

Other problems may be caused by defects in the design and the examples have been given of defects in the site and sample numbering. These problems could be solved by redesign of the database if staff resources are available.

### 5.1 DATA PROBLEMS

1. Same site numbers being given to samples collected at different locations. This has been discussed in an earlier section. Sample and site renumbering is not documented in the database.
2. Incomplete data in the database so that sample analyses may not have locations or vice versa. This is mainly a problem in the MRP and is discussed and quantified in Hobbs et al. (2000).
3. Little validation or checking of data loaded to the database.
4. Original source data not loaded by G-BASE only the modified or 'normalised' data.
5. Data dispersal due to reduction in storage space and source with the original source data being moved to the BGS data archive before entry into the database.

### 5.2 DESIGN PROBLEMS

1. External data cannot be included in the present system because it relies in information like Batch number to enforce the referential integrity. Data in external publications rarely contains sufficient information about the analytical methods and detection limits to comply with the database standards.
2. BGS environmental data cannot be loaded to the database because it does not comply with the numbering system.

### 5.3 STAFF PROBLEMS

1. Reductions in the level of staffing have led to an over-reliance on a few key staff (4) who are difficult to replace if they move to new projects to enhance their career prospects or retire. Replacements should be trained before these staff move.
2. Loss of expertise due to poor documentation of QA/QC procedures.
3. Documents and records kept by G-BASE not systematically reported as internal reports.
4. Ignorance of the database structure and consequent failure to load QA/QC information
5. Loss of DTI contract and reduction in minerals related funding.

## 6 Conclusions and Recommendations

The BGS Geochemistry Database has reached a critical stage in its evolution with the reduction in funding due to the loss of the DTI contract, the retirement of the author and the diversion of two key personnel to other projects. It is important that the BGS management address this problem and train younger members of staff to take over these responsibilities. The problem has been predictable and has been flagged up when staff have been diverted to overseas projects such as Morocco and Mauritania. Reduction in funding or changes in personnel will cause problem but this is the job of the programme managers and HODs to find and train suitable replacements.

This problem has been exacerbated by of the lack of quality assurance procedures in previous years and poor documentation of procedures such as data levelling. The Geochemical Database was well documented in the early years and probably the large number of design reports hid the fact that the data reporting procedures were very lax. Database developers documented their work very well but were unconcerned about the quality of the data being entered or its completeness. The MRP staff were always better at collecting samples and analytical data rather than archiving and preserving it. There are several instances where MRP projects were aborted because of poor results and these were never written up or archived in any way. It is likely that this data has been lost. With the reduction in funding this should have increased the need to look after the data rather than reduce the data management.

The G-BASE decision to store 'normalised' data in the database made this situation worse as the original sources were lost in many cases. The MRP made the decision to store only original data and any secondary data such as gridded, log transformed or principal components data was calculated as needed and stored in project files rather than the database. Changes in the reporting procedures in the BGS analytical laboratories also resulted in data loss and in the failure to collect important metadata.

External laboratory analyses always caused problems because it came directly to the project geologist rather than through the structured reporting system of the BGS laboratory. Batches were amalgamated, method of analysis not recorded, dates missed off data compilations. All these caused problems. When one laboratory, such as ACME, was used for a long period of time the data management tasks became easier and most of the data and metadata has been captured. This problem is also very important where geologists have collected significant batches of results on particular rocks such as limestones. Batches may have been assembled over several years and metadata has been lost. Data is probably still residing in filing cabinets and not in digital format.

The costing of the staff time necessary for data management and administration - managing, checking and validating data in the database - is a necessary overhead of the database and without it the database will become increasingly corrupt and fall into disuse. How much of this should be paid for by corporate funding from the Information Management Programme and how much from project funding from EMGB is open to debate, but this problem has got worse with the loss of CR money from the Department of Trade and Industry. There is a strong case for the transfer of some funds back from IMP to EMGB, thereby partially reverting to the system existing before the introduction of matrix management.

Notwithstanding all these problems the Geochemistry Database has been successful and, despite operating with a very small staff, has managed and delivered geochemical data to G-BASE, the MRP and its successor the Minerals Programme. The multidisciplinary and mineral potential studies undertaken in recent years would not have been possible without it. Without the database it would not have been possible to prepare reasonably consistent geochemical atlases or maps of the United Kingdom. The forethought of the geochemists in designing data cards and structures in 1970 also greatly assisted this process.



## **6.1 FUTURE DEVELOPMENTS AND RECOMMENDATIONS**

### **6.1.1 Data Warehouse**

A data warehouse is an un-normalised copy of a relational database configured in a single large table for rapid retrieval of whole records. It is commonly used to answer external queries, which just require the retrieval of whole rows rather than complex queries involving several table joins. The G-BASE data with just analytical results and locations could be configured as a single very large table, which could be available to BGS and external users. Simple spatial queries from internet browsers or GIS packages could be answered from the data warehouse.

The advantage of this approach is that details of the relational database structure are hidden from the user and retrievals are quicker because they don't involve multi-table joins. The disadvantages are that you are working on a copy of the data, which may not have been updated, and it cannot answer more complex questions.

It is recommended that a feasibility study is prepared by Alan Mackenzie on the creation of a data warehouse in ORACLE.

### **6.1.2 Expansion to a wider variety of BGS users**

The data warehouse approach may be suitable for the average BGS user who will probably lack the database skills to use the relational database. Data is already made available to BGS users through the GDI and internal ArcView gdi so a data warehouse may not add extra functionality. The SQL script that prepares data for the GDI effectively creates a data warehouse as a *.dbf* file in ArcView. The problem is probably not to do with data access but rather advertising to BGS staff what data is available. Steps have been taken to solve this problem and Johnson and others (2004) have written a report for the SIGMA project in which geochemical data is a recommended reference dataset.

It is recommended that a publicity campaign is mounted to advertise the Geochemistry Database and its data availability to BGS staff and other potential users.

### **6.1.3 Unified management of the MRP and G-BASE data**

With the removal of DTI funding there is no staff effort for entering any of the old MRP data to the database and ISP resources may also be reduced. Any external queries on the MRP data will have to be funded out of enquiries or charged back to the enquirer. To maximise the use of available skilled staff it is recommended that there is much closer integration of G-BASE and former Minerals staff. There will need to be further training for the database and data management team to familiarise themselves with the totality of the data.

It is recommended that the management of the Geochemistry Database is integrated into one team.

### **6.1.4 Better training in QA and DMP for G-BASE staff**

It has become apparent that detailed knowledge of the database and its data has been restricted to just two people, both of whom are busy with other tasks. Whilst QA/QC procedures are in operation in the field and in the laboratory these do not seem to have been followed in the data manipulation and administration. Many data procedures such as data 'levelling' or 'normalisation' were carried out without keeping records and modified data was stored back into the database in place of the original. Detailed Quality Assurance and Data Management Plans should be in place to guide new users in the procedures to be followed. External training may be required to implement this.

It is recommended that G-BASE and Minerals staff receive further training in QA methods and in creating a Data Management Plan.

#### **6.1.5 The role of the Data Administrator**

The role of the Data Administrator is very undervalued in BGS but without a careful, diligent and skilled person to manage data in the database its quality deteriorates. This role is distinct from that of the database developer, who may only work for a few weeks creating the database, or the database manager who understands both the data and the database but rarely has time to manage it in any detail. The role of the data administrator is to manage the quality of the data, make corrections, to ensure data integrity and backup, and to assist users in data retrievals.

It is recommended that the role of Data Administrator is given extra recognition in BGS.

#### **6.1.6 Database developers trained in EMGB methods**

With full implementation of the matrix management system in BGS the management of the database developers (IS HOD) has been removed from that of the data collectors (EMGB) and also from the database management (IMP). This potential three way split has some unfortunate consequences with data collectors lacking database skills to properly administer the data, data administrators being downgraded because they don't fit easily into the matrix and the database project leader possibly not understanding either the database or the data. Whilst the day of the generalist database developer, data collector and project leader all rolled into one is probably over there is a need for the database developer to be better trained and conversant with the procedures in the EMGB. Database developers should be expected to load and extract real data before they hand over the finished project. They also should provide detailed user manuals and train staff in use of the database.

It is recommended that database developers become more involved in the database work within EMGB.

# Glossary

This glossary is largely taken from the CASE\*Method reference guide published by Oracle with some simplification and additions.

*Application System* A name given to a collection of business functions, entities, programs and tables.

*Arc* A means of identifying two or more mutually exclusive relationships.

*Attribute* Any detail that serves to qualify, identify, classify, quantity or express the state of an entity: any description of 'a thing of significance'.

*CASE* Computer-Aided Systems Engineering is a combination of graphical, dictionary, generator and other software tools to assist in the computer development of relational databases.

*Character* The format of an attribute, which may contain alphabetic characters or numeric digits.

*Column* A data field which implements an attribute.

*Database* A structured collection of tables or files under the control of a database management system.

*Dataflow* A named flow of data between business functions.

*DBMS* A database management system, normally a computer program or system that is used to manipulate and manage data, and to manage user access in a multi-user environment.

*Domain* A set of business validation rules that apply to a group of attributes. For example, a list of values, a range, a qualified list or range or a combination of these.

*Entity* A thing of significance, real or imagined, about which information needs to be known or held.

*Entity – Relationship Diagram* A diagram that depicts entities, the relationships between them and the attributes used to describe them. Entity – relationship model is a synonym.

*Field* A means of identifying an item of data within a file. It can be character, date or number in format.

*File* A method of implementing part or all of a database.

*Foreign key* One or more columns in a table that implement a many to one relationship that the table in question has with another table.

*Format* The type of data that an attribute or column represents. For example, character, date, number etc.

*Index* A means of accessing one or more rows in a table with the aim of increasing performance or enforcing uniqueness.

*Key* Any set of columns that is frequently used for retrieval of rows from a table.

*Metadata* Data that describes other data, which may describe or summarise data in one or many tables. Discovery metadata enables to user to identify the contents of one or more tables. Technical metadata describes the arrangement and format of the columns or attributes.

*Normalisation* A step by step process that produces either entity or table definitions that have no repeating groups, the same kind of values assigned to attribute or columns, a distinct name, and distinct and uniquely identifiable rows.

**Null** A data item that has no current value or the value is unknown.

**Primary Key** A set of mandatory columns in a table that is used to enforce uniqueness of rows.

**RDBMS** A relational database management system.

**Record** A record in a table that consists of values for each attribute or field.

**Relationship** Any significant way in which two entities of the same or different type may be associated.

**Row** An entry in a file or table that consists of values for each column. A spreadsheet has *rows*.

**SQL** Structured Query Language. The standard language for relational systems covering not only queries but also data definition, manipulation and integrity.

**Sub-entity** A sub-type of entity. An entity may be split into two or more sub-types, each of which has common attributes and relationships.

**Table** A tabular view of data used in a relational database to hold one or more fields of data. It is often an implementation of an entity. Tables are the logical and perceived data structure, not the physical data structure, in a relational system.

**Unique identifier** Any combination of attributes that serves to uniquely identify an occurrence of an entity.

## 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.

COATS, J.S. AND HARRIS, J.R. 1995. Database design in geochemistry: BGS experience. *In Geological Data Management* (ed. Giles J R A), Geological Society Special Publication No 97, 25-32.

COATS, J.S., HOBBS, S.F. AND MACKENZIE, A.C. 1996. The user interface: a critical factor in the successful use of geochemical databases [abstracts]: *In: Applied geosciences*; Geological Society 1996; p.44.

GALLGHER, M.J., MICHIE, U.MCL., SMITH, R.T. AND HAYNES, L. 1971. New Evidence of uranium and other mineralization in Scotland. *Transactions of the Institution of Mining and Metallurgy, Section B Applied Earth Sciences*, 80B, 140-171.

HARRIS, J.R. AND COATS, J.S. 1992. Geochemistry Database: data analysis and proposed design. *BGS Mineral Reconnaissance Programme Report, (1992), No.125, 93pp.*

HARRIS, J.R., GLAVES, H. AND COATS, J.S. 1992. Geochemistry Database report 2: A proposed integrated sample numbering system. *BGS Technical Report WP/92/5R Applied Geochemistry Series.*

HARRIS, J.R., KHAN, A.H., STEPHENSON P. AND WILD, S.B., 1992. Geochemistry Database report 3: Guidelines for the detailed physical design, implementation and management. *BGS Technical Report WP/92/6R Applied Geochemistry Series.*

HARRIS, J.R., COATS, J.S., WILD, S.B., KHAN, A.H., AND STEPHENSON, P. 1992. Geochemistry Database report 4: The User Requirement. *BGS Technical Report WP/92/12R Applied Geochemistry Series.*

HARRIS, J.R., NICHOLSON, C.J., AND COATS, J.S. 1993. Geochemistry Database report 6: Standardisation of geochemical field cards 1970-1992. *BGS Technical Report WP/93/20R Applied Geochemistry Series.*

HARRIS, J.R., KHAN, A.H., WILD, S.B., MACKENZIE, A.C., AND COATS, J.S. 1994. Geochemistry Database report 5: Detailed physical design.. *BGS Technical Report WP/94/15R Applied Geochemistry Series.*

- HAWKES, H.E. AND WEBB, J.S. 1962. *Geochemistry in Mineral Exploration*. Harper and Row, New York, N. Y., 415pp.
- HOBBS, S.F., COLLINS, G.L. AND COATS, J.S. 1998. *Geochemistry Database report: The use of data loading programs on desktop PCs*. *BGS Technical Report WF/97/12R*, 1998, 38pp.
- HOBBS, S.F., KIMBELL, S.F., COATS, J.S. AND FORTEY, N.J. 2000. *BGS databases for mineral exploration: status in 2000*. BGS Research Report RR/00/11. DTI Minerals Programme Publication No. 6, 2000, 55pp.
- INSTITUTE OF GEOLOGICAL SCIENCES (IGS). 1978a. *Regional Geochemical Atlas: Shetland*. (London: Institute of Geological Sciences.) ISBN 0 85272 061 0
- INSTITUTE OF GEOLOGICAL SCIENCES (IGS). 1978b. *Regional Geochemical Atlas: Orkney*. (London: Institute of Geological Sciences.) ISBN 0 85272 060 2
- INSTITUTE OF GEOLOGICAL SCIENCES (IGS). 1979. *Regional Geochemical Atlas: South Orkney and Caithness*. (London: Institute of Geological Sciences.) ISBN 0 85272 064 5
- JOHNSON, C C, COATS, J S, BREWARD, N, ANDER, E L, and MACKENZIE, A C. 2004. *Geochemical data as a standard reference data set for the SIGMA project*. *British Geological Survey*, BGS Internal Report Number IR/04/026N.
- LEAKE, R.C. AND SMITH, R.T. 1975. A comparison of stream sediment sampling methods in parts of Great Britain. *In Geochemical Exploration, 1974*, (eds. Elliott, I.L. and Fletcher, W.L.), Elsevier, Amsterdam. 577-594.
- MACKENZIE, A.C. 1997. *Geochemistry Database report 9: Data retrieval system*. *BGS Technical Report WP/97/2R Applied Geochemistry Series*, 1997.
- MACKENZIE, A.C. AND JOHNSON C.C. 1996. *A data retrieval interface for the Geochemistry Database using Microsoft Access 7.0*. *BGS Technical Report WP/96/16 Applied Geochemistry Series*, 1996.
- PATEL, D. AND MACKENZIE, A.C. 1994. *Geochemistry Database report 8: Data loading sub-system*. *BGS Technical Report WP/96/12R Applied Geochemistry Series*, 1996.
- PATEL, D. AND MACKENZIE, A.C. 1994. *Geochemistry Database report 11: User Manual (development version)*. *BGS Technical Report WP/96/16 Applied Geochemistry Series*, 1996.
- PLANT, J. 1971. Orientation studies on stream sediment sampling for a regional geochemical survey in Northern Scotland. *Transactions of the Institution of Mining and Metallurgy, Section B Applied Earth Sciences*, 80B, 324-344.