British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Information Management Focus and Operations: Delivery Plan 2010 - 2011 and Beyond

Information Management Programme

Internal Report OR/11/033

BRITISH GEOLOGICAL SURVEY

INFORMATION MANAGEMENT PROGRAMME

INTERNAL REPORT OR/11/033

Information Management Focus and Operations: Delivery Plan 2010 - 2011 and Beyond

L Ault, P H O Henni, S F Hobbs, R R Luckett, T McCormick, A A McKenzie and H M Glaves

Contributors:

C E Morley, W J Rowley, C P Royles, R P McIntosh

Keywords

Report; keywords: Information Management, Data Standards, Data Curation.

Bibliographical reference

L Ault, P H O Henni, S F Hobbs, R R Luckett, T McCormick, A A McKeenzie and H M Glaves. 2011. . *British Geological Survey Internal Report*, OR/11/033. 46pp.

Keyworth, Nottingham British Geological Survey 2011

BRITISH GEOLOGICAL SURVEY

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

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

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

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

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

British Geological Survey offices

BGS Central Enquiries Desk

Tel 0115 936 3143 Fax 0115 936 3276

email enquiries@bgs.ac.uk

Kingsley Dunham Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3241 Fax 0115 936 3488

email sales@bgs.ac.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

Tel 0131 667 1000 Fax 0131 668 2683

email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090 Fax 020 7584 8270

Tel 020 7942 5344/45 email bgslondon@bgs.ac.uk

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

Tel 029 2052 1962 Fax 029 2052 1963

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800 Fax 01491 692345

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

Tel 028 9038 8462 Fax 028 9038 8461

www.bgs.ac.uk/gsni/

Parent Body

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

Tel 01793 411500 Fax 01793 411501

www.nerc.ac.uk

Website www.bgs.ac.uk

Shop online at www.geologyshop.com

Preface

The strategic context of information management has changed over the past two years in a number of ways:

- Budgets are declining due to the government's response to the economic down turn;
- NERC has developed and agreed its first Science Information Strategy which is being implemented between 2010 and 2014 (see: www.nerc.ac.uk/research/sites/data/sis.asp); and
- as part of the Science Information Strategy implementation NERC has published a new Data Policy (see: www.nerc.ac.uk/research/sites/data/policy2011.asp)

Like all parts of NERC information management in BGS will see a decline in budgets. The National Capability Advisory Group for Earth Sciences has proposed a 13% decline in funding over the next three years for the National Geoscience Data Centre. There may be other cuts on top of this decline. Our response is:

- to work more efficiently by seeking saving in work flows;
- to work closely with BGS scientists to focus resources on their science priorities;
- to work with others, both in NERC and outside NERC, to develop methods and approaches to information management in collaboration rather than on our own; and
- to accept that backlogs of unprocessed data, which have been declining over the past decade, will once again start growing.

The Science Information Strategy seeks to create a clear distinction between the information management programmes of NERC centres and the NERC Environmental Data Centres such as the National Geoscience Data Centre. In some parts of NERC this distinction is clear as in the case of the British Oceanographic Data Centre which is closely associated with the National Oceanographic Data Centre, but has a distinct funding stream directly from NERC. In the case of BGS and the National Geoscience Data Centre there are currently no clear boundaries between the bodies.

The next few years will require NERC and BGS to agree what the nature of these boundaries are, what components of the existing BGS Information Management Programme will be classified as being part of the National Geoscience Data Centre and what components remain in the BGS internal information activity. This will be a challenging discussion and the outcome will inevitably appear arbitrary. However, the work has to be done to comply with the NERC Science Information Strategy.

The revised NERC Data Policy places more rigorous requirements on NERC funded scientists, both those in NERC centres and university academics, to deposit datasets when their research project concludes. The policy requires:

- all environmental data of long-term value, produced by the activities funded by NERC to be made openly available for others to use;
- data should be well managed during the life of the project and data of long-term value should be deposited in a NERC Environmental Data Centre;
- data management plans will be produced by all NERC funded scientists at the commencement of the project;

- close working between the NERC Environmental Data Centres and the NERC funded scientists to ensure that datasets are delivered in the appropriate formats and with adequate metadata at the end of the projects; and
- those funded by NERC who do not comply with the requirements of the data policy may have payments withheld or may become ineligible for future funding from NERC.

Over the next few years information management in BGS faces a number of challenges which will change the way we work. There are real benefits to the changes proposed, but given the current economic climate they will be hard to achieve.

Jeremy Giles

Head of Information Management

May 2011

Foreword

This report provides an overview of the tasks undertaken by the Information Management Focus and Operations Team of the BGS for the year 2010 – 2011 and includes a forward look to future years.

Acknowledgements

A large number of individuals in BGS have contributed to work described here. Acknowledgement is made to the contributions and hard work of all the staff who worked in the Information Management Focus and Operations Team during 2010 – 2011, to all those who contributed via strategic deliverables and to all data managers who realised the value of their data and took pains to organise and secure it for future users.

Contents

FIGURES

PLATES

TABLES

Summary

This report provides an overview of the work undertaken by the Information Management Focus and Operations Team for the financial year 2010-11 and includes a forward look for future years.

Currently the Team's projects cover specialist data management activities under five geoscience themes, geophysics, geology, hydrogeology, geochemistry and marine plus a multi tasked project covering corporate metadata, proactive data verification and project data storage. The Geo-Seas project, a Pan-European e-infrastructure project for the management of marine geological and geophysical data, is also managed under this Team.

1 Introduction

The Information Focus and Operations (IMFO) Team have the broad remit of promoting good data management within BGS. The Team has responsibility for BGS Corporate metadata records, assuring project information assets and Storage Area Network (SAN) resources; management of Data Management Plans and improving data in corporate databases by proactive data verification activities. In addition, through the Information Management (IM) Sector Managers for the primary geoscience disciplines they ensure that data are properly managed and preserved and can be used and potentially re-used by future researchers, exploited commercially and educationally. The IM Sector Managers are responsible for the maintenance and management of corporate datasets for their discipline.

A vital function of this team is to ensure that good quality data produced as the result of research activities is stored in secure conditions for the long-term and is at the same time made findable and readily accessible for future users.

In addition IMFO oversees and provides the co-funding for Geo-Seas a Pan-European e-infrastructure project for the management of marine geological and geophysical data (FP7-Infrastructures-2008-1.2.2).

2 Overview of Activities

Projects and tasks undertaken by the IMFO Team are:

1. Metadata, Data Management and Audit, including:
    a. Management of SAN resources.
    b. Corporate Metadata
    c. Proactive Data Verification
    d. Information Management Training
2. Geophysical Data Management
3. Geological Data Management
4. Hydrogeological Data Management
5. Geochemical Data Management
6. Marine and coastal Data Management
7. Geo-Seas
8. Strategic Deliverables

3 BGS Wide Activities

3.1 METADATA, DATA MANAGEMENT AND AUDIT

Linda Ault

3.1.1 Data Management Planning

It is NERC Policy that:

"all environmental data of long-term value generated through NERC-funded activities must be submitted to NERC for long-term management and dissemination."

The Data Management Planning System (DMPS) has been designed to record the information necessary for project data management. It covers such things as data types, datasets used, quality assurance applied, Intellectual Property Rights (IPR) and confidentiality issues. It is also the main tool for the management of the Storage Area Network (SAN) resources.

3.1.2 Training

A training course in Data Management in BGS was provided for the BGS Internal Auditors in November 2010. This promotes better policing of data management procedures, such as Project Data Management Planning, long-term asset storage, project closure and subsequent release of SAN space for re-use by active projects.

3.1.3 Data Management Planning System

Since the shutdown of the BGS Project Management System (PMS) less functionality has been available in the DMPS since it is not possible to link systems directly into the NERC wide Resource Management System (RMS). The Information Management Database Solutions Team (DBS) have worked in the background to provide linking functionality to keep local systems working. A revamp of the DMPS was carried out last year in order to keep the system viable. This year the system has been updated to reinstate some lost functionality and to improve user-friendliness. Improvements have included:

3.1.3.1 EDITING

It is now possible to edit Data Management Plans where previously it was possible only to save an unchangeable version.

3.1.3.2 CLARIFICATION OF THE FORMS

Form labels and explanatory texts have been refreshed.

The layout on some pages has been simplified, for example the page listing the summary of SAN allocations (project space and quota changes) for a project.

Some input fields on the form were reportedly too small in the previous version, they have now been increased.

Hyperlinks between pages in the system have been added or made obvious through the use of underlined blue lettering.

3.1.3.3 IMPROVEMENTS TO THE SEARCH FUNCTION

More database fields are used in the searches, widening the range of results returned.

More ordering options have been added to the search results, i.e. order by Title or Most Recent.

Additional information is now displayed in the search results, for example where a project exists in the PMS, a Bid ID will be shown.

3.1.3.4 REVISION LOG

A new history page has been added showing when and by whom any changes to the DMP have been made.

3.1.3.5 Information Management Contacts and Help for the DMPS

An intranet page was created to coincide with the launch of the updated DMPS. This page provides contact details for the IM SAN Management team and senior data specialists. There are also links to guidance pages and general information about the SAN. The help page is linked into the DMPS home page and is also available via the intranet Project Management menu.

3.1.3.6 Outstanding issues

Resources this year did not allow for alerts to Records and Collections staff (when projects are initiated) to be reinstated.

3.1.3.7 Workarounds

The demise of the PMS means that IMFO staff are not alerted when new projects begin or active projects end. To assist with this, Business and Operations Support staff have instigated manual notifications to IMFO. By querying the RMS on a monthly basis they are able to supply IMFO staff with lists of projects starting up or closing. IMFO staff then contact relevant Project leaders to alert them to create a DMP or to ensure that significant project data are stored securely and accessibly for the long term.

It is planned that automatic notifications will be structured into the next update of the RMS project workflow.

3.1.4 Management of the SAN

A SAN is a high-speed network dedicated to information management and it consists of various levels of storage hierarchy that permit large volumes of data to be stored in a secure manner. The SAN allows for consolidation of storage resources, centralised management of these and greater opportunities for day-to-day maintenance of data; e.g. improved data backup. IMFO project staff work together with Systems and Network Support (SNS) in the management of SAN resources. SNS are responsible for ensuring that at the technical level all aspects of the SAN infrastructure are fully operational.

Projects are allocated space on the secure corporate data holding system, the SAN for the lifetime of the project. The information recorded in the DMPS enables the identification of such assets and helps monitor usage of SAN space, a necessary requirement for efficient space usage. It is a corporate aim that at project closure digital assets that have long term value will be entered into the corporate records of the NGDC.

The SAN Management Committee is responsible for overseeing the management of the SANs in Keyworth, Murchison House and Wallingford, for implementing SAN management policy and providing information on SAN issues to the Information Systems Steering Committee (ISSC). The SAN Management Committee includes representatives from Records Management, SNS and the Head of Information Management and is chaired by the IMFO Team Leader.

3.1.5 The Digital Archive

Working data are held on the spinning disks of the SAN and when a project ends or a dataset is superseded the requirement for speedy access to the data may no longer exist. However, the information may need to be stored for reference purposes or for evidence of provenance. The Digital Archive was developed to store such information. The application enables a data holding to be archived to tape while retaining metadata describing the data to be held in an Oracle™ database. The metadata can be searched to find specific records of interest which can then be retrieved to disk (for access) as required.

The Digital Archive application was developed by a collaboration of the Records, IMFO and Database Solutions Teams; the Database Solutions Team providing the coding and database support to the system requirements as defined by Records and IMFO. The original system, released October 2007 has been updated in 2010/11 to take account of user feedback, BGS restructuring, software upgrades and two extra operational requirements:

1. **The Vault.** This is a special area of the SAN where data from closed projects can be stored for the short term, allowing data to be referenced speedily for any issues concerning customer feedback or for reference when setting up follow on projects.
2. **Archive Review.** This feature of the updated Digital Archive allows Records' staff to review archives for continuing relevance or for updates to confidentiality/access restrictions.

The opportunity is being taken to provide a linked web page containing help information and contact details for Records and IMFO staff who can provide guidance and support. It is envisaged that the offer of training courses will also feature in the relaunch plans.

3.1.6 Corporate Metadata

BGS and the NGDC need to have a good understanding of the value of the data they hold. They need to advertise the data and make them available for others to make use of them. To quote the NERC data policy:

NERC has a policy on data in order to:

a. Ensure the continuing availability of environmental data of long-term value for research, teaching, and for wider exploitation for the public good, by individuals, government, business and other organisations.

b. Support the integrity, transparency and openness of the research it supports.

c. Help in the formal publication of data sets, as well as enabling the tracking of their usage to be tracked through citation and data licences.

d. Meet relevant legislation and government guidance on the management and distribution of environmental information.

Discovery metadata records describe datasets in a way that enables potential users to be able to identify whether or not the data will be of use to them, as such it is a vital tool in carrying out the policy outlined above. BGS currently stores discovery metadata, conforming to the international standard BS ISO 19115:2003, Geographic information - Metadata, describing each data holding. The metadata records are held in a dedicated database in Oracle™.

3.1.6.1 METADATA MAINTENANCE

It is one of the tasks of the IMFO Team to ensure that the metadata records are up to date and that new records are created for new datasets. Project staff liaise with the other NERC data centres to ensure conformity in standards across the research council. They work with the Database Solutions Team and Web Systems to facilitate the development of mechanisms to allow the formal publication of data sets. Currently distribution is via the NERC Data Discovery Service.

3.1.6.2 THE INSPIRE DIRECTIVE

Relevant metadata records must also be checked for compliance with the INSPIRE Directive. This directive aims to make consistent information on spatial data sets about the environment available and to create services for accessing these datasets, enabling them to be more easily shared or combined, benefitting the development and monitoring of environmental policy and practice in all Member States and across the European Community. Under this directive

metadata for spatial datasets corresponding to Annex I and II should be made available by 3rd December 2010.

Table 1 Scope of the INSPIRE Directive

This Directive shall cover identifiable collections of spatial data, hereinafter "spatial data sets", which fulfil the following conditions: (a) they are related to an area under the jurisdiction of a Member State or to its exclusive economic zone/search and rescue region, or equivalent; (b) they are in electronic format; (c) they are in the possession of either of the following: (i) a public authority, having been produced or received by a public authority, or being managed or updated by that authority; (ii) a natural or legal person on behalf of a public authority; (iii) a third party to whom upload services have been made available in accordance with Article 17(3); (d) they relate to one or more of the themes listed in Annexes I, II or III.

Annex I	**Annex II**
Coordinate Reference Systems	**Elevation**
Geographical Grid Systems	**Land cover**
Geographical Names	**Orthoimagery**
Hydrography	**Geology**

It is the responsibility of the IMFO Team to collect, collate and record INSPIRE compliant metadata into the corporate system. The initial selection process selected 120 datasets within the scope for INSPIRE with new datasets added as they are created.

3.1.6.3 DISCOVERY METADATA HOLDINGS

Currently there are ~ 474 discovery metadata records in the system. These records describe data holdings as diverse as microfossil collections, major corporate databases, hardcopy records and individual datasets. Only datasets with up to date compliant metadata are marked as "complete" by the metadata manager and it is only these datasets that are advertised outside the organisation.

Datasets with poor metadata or with little contact information (e.g. the person responsible for the dataset has moved on) are to be flagged as "required", defined in ISO 19115, as "data needs to be generated or updated".

Discontinued datasets (merged into new datasets or superseded by newer systems) are to be flagged as "obsolete" (ISO definition: "data is no longer relevant").

Older versions of datasets are to be flagged as "historicalArchive" (ISO definition: "data has been stored in an offline storage facility").

There is no restriction on datasets listed as "obsolete" or "historicalArchive" having metadata listed as "complete" and therefore being available across the web. Datasets listed as "required" would be listed internally only until sufficient information is gathered on them to locate them and identify their condition and usefulness.

3.1.6.4 DATASETS AND THE SCIENCE INFORMATION STRATEGY

The NERC Data Policy is being reviewed and revised as part of the work for the Science Information Strategy implementation programme. Changes to the Data Policy "reflect a change in approach to how NERC will identify data of long-term value and the data management requirements it places on NERC funded researchers and the need for clarity in how NERC will charge for Data and for Information Products".

The need to classify BGS datasets into Data or Product became apparent. As a first pass the discovery metadata is being used to help decide which category datasets fall into. The case is not always clear cut; there are more things to consider besides data/product. There are confidentiality and IPR issues, free products and combined data/product datasets. A starting point is to make an initial assessment and provide a reason for the decision made. This work is being carried out by the Information Management Sector Managers in conjunction with metadata project staff.

3.1.6.5 TECHNICAL METADATA

Technical metadata are described as "the technical aspect of the metadata describing how the data within the dataset are stored within a digital database. It also covers such aspects as the contents of list domains and the maximum and minimum values of the range domains".

These metadata record details of the Oracle™ tables which make up specific databases. The table contents are described and responsibilities for upkeep defined. Work is underway to link technical metadata definitions to each discovery metadata entry covering an Oracle™ database.

A system of "alerts" has been set up within the database management system to inform the technical metadata manager of missing metadata so that such records can be updated.

3.1.6.6 OTHER ISO 19115 METADATA COLLECTIONS

The database designed to hold basic Discovery metadata for BGS datasets is also used to house other metadata collections. These collections are managed by the specific project or science area they belong to. All these collections use the core metadata elements, those mandatory under ISO19115 but may include extra elements relevant only to their specific type of data, e.g. metadata for 3D modelling contains extra elements to record parameters relevant specifically to models.

2010 saw the launch of the Earth Science Academic Archive metadata. This metadata holds information on projects and data received into the National Geoscience Data Centre from academics. These records and the metadata for them is managed by the BGS Records Management Team as part of the NGDC record holdings and the metadata are a specific collection within the ISO 19115 metadata tables in Oracle™. Many of these datasets will fall under the requirements for the INSPIRE directive and BGS Teams (Records, IMFO, Database Solutions, Web) will be working together to ensure that relevant metadata from appropriate collections are made available via the NERC Data Discovery Service (DDS).

3.1.7 Verification Activities

Claire Morley

3.1.7.1 GEOSCIENCE INFORMATION ISSUES LOG

The Geoscience Information Issues Log provides the means for internal BGS users to record any problems with BGS data. The Issues Log is accessed via the Intranet Data Access (IDA). Any problems reported by users are recorded in a dedicated Oracle™ database. Most current issues relate to problems with the data held in the Single Onshore Borehole Index (SOBI) or with the associated borehole scans. However, issues regarding the borehole geology database, Lexicon, GIS layers and Site Investigation reports have also been recorded.

3.1.7.2 INITIAL VERIFICATION WORK

Prior to the setting up of the current Verification section in August 2008 a backlog of issues (amounting to about 1000 issues) had built up and the issues log table was difficult to manage. Sorting out the backlog involved categorising each issue by its type. This enables the issues to be sorted and passed to the relevant data managers who can deal with the problem.

Correction procedures are agreed for each type of problem. The procedures are documented as flowcharts; ensuring consistency of approach and maintaining a record of current methodology.

There are currently 26 issue categories:

Table 2 Issues Log Categories

Issue code	Description of issue	Number of occurrence on the issues log (Jan 2011)	Number of resolved issues (Jan 2011)	Percentage of issues resolved (Jan 2011)
WLWIWL	Wallingford Wrong Images in WellMaster	295	265	90%
CRRCTD	Corrected	283	283	100%
KWSA	Keyworth Wrong Scans Associated	220	205	93%
SUPREC	Duplicate Record	195	17	9%
SBNGR	SOBI Missing/Wrong Grid Reference	113	75	66%
KWMSP	Keyworth Missing Scanned Page	106	99	93%
KWPQS	Keyworth Poor Quality Scan	94	89	95%
SBOTH	SOBI Other	77	44	57%
INDSHT	Index Sheet	69	69	100%
OTHER	Other	64	53	83%
KWBP	Keyworth Problem with Batch of Scans	57	46	81%
BI-SNI	Bad Image/Scan Not Indexed	30	22	73%
SBDRLN	SOBI Missing/Wrong Drilled Length	26	26	100%
SBNAME	SOBI Missing/Wrong Borename	24	22	92%
WLMSP	Wallingford Missing Scanned Page	24	23	96%
KWMB	Keyworth Multiple Bores under One Record	22	0	0%
WLOTH	Wallingford/WellMaster Other	19	13	68%
?	We have not assigned a value yet (and it might not be applicable)	17	8	47%
WLSBWL	Wallingford Poor SOBI-WellMaster Match	15	3	20%
MHMSP	Edinburgh Missing Scanned Page	8	6	75%
WLPQS	Wallingford Poor Quality Scan	7	7	100%
ADJSHT	Adjustment Sheet	5	5	100%
GIS	GIS Problem	5	3	60%
SCNOTH	Scan Other	4	2	50%
WLBP	Wallingford Problem with Batch of Scans	4	3	75%
MHBP	Edinburgh Problem with Batch of Scans	1	1	100%

3.1.7.3 ISSUES LOG TABLE REDESIGN

Limitations in the table structures held in Oracle™ made the issues log difficult to manage. A redesign was needed in order to track the progress of issues more easily. Additional fields were required to record information such as the issue category, the status of the issue and any notes on the issue. To accommodate these new management requirements the issues log was redesigned in April 2010 in collaboration with database development staff.

The Issues Log is accessed and used in the same way as before but information can now be recorded about the type of issue, the stage the issues is at, how it has been corrected and any comments. The new updated application includes email notifications to the user who reported the issue stating how the issue has been resolved or explaining any outstanding problems.

3.1.7.4 FURTHER VERIFICATION WORK

Other verifications tasks include:

1. Comparing SOBI to other datasets and ensuring consistency of data across all systems by amending values as necessary.
2. Correcting automatically indentified errors in Borehole Geology. Since January 2009, 50% of the errors have been cleared, reducing the errors from just over 90,000 to 45,500.

Further details on verification activities including details of specific issues and their resolution are to form the basis of a technical report (in preparation).

4 Geoscience Themed Activities

4.1 INFORMATION MANAGEMENT SECTOR MANAGERS – SENIOR DATA MANAGERS

Currently data management activities are divided into five sectors, each sector being the responsibility of the senior data manager for that area of science.

Table 3 Information Management Sector Managers

Sector	Manager	Manager's Location
Geophysics	Richard Luckett	Edinburgh
Geology	Tim McCormick	Keyworth
Hydrogeology	Andy McKenzie	Wallingford
Geochemistry	Susan Hobbs	Keyworth
Marine Geology	Paul Henni	Edinburgh

Although nominally representing their area of science, the Sector Managers will also provide local data management support for scientists at their office location as required.

4.1.1 Main Activities of IM Sector Managers

A delivery plan for the management and support of IM activities produced in consultation with BGS Team Leaders – this document.

Key corporate database content enhanced and maintained: e.g. Lexicon, Geochemistry etc.

Selected dataset specific management procedures developed, emplaced and operating.

Sector Manager duties include:

1. Advising Project Leaders on Information Management.
2. Ensuring projects produce adequate Data Management Plans.
3. Being the second signatory on Data Management Plans. (“Checked by”)
4. Ensuring dataset producers provide Discovery Metadata.
5. Advising on data retention and disposal at project closure (see Information Management procedure 1.1.2 Project Asset Assessment and Management)

One of the important tasks for the Sector Managers is to meet with BGS Team Leaders to get a feel for the kinds of data being produced/collected, what the issues are, what future storage requirements/systems will be needed.

To this end, the BGS Teams have been assigned to a specific IM Sector Manager in the first instance. Some BGS Teams will have overlaps between more than one IM sector but there is a nominal first contact for each Team, this contact will be the Sector Manager who checks relevant Project DMPs and liaises with the Team Leader. This year, Team responsibilities have been mapped as follows:

Table 4 Sector Managers’ Team Responsibilities

4.1.1.1 GEOPHYSICS

Earthquake Seismology
Geomagnetism
Volcanology
CO2 Storage
Advanced Seismic Techniques
Geophysical Tomography

4.1.1.2 GEOLOGY

Palaeoclimate & Palaeoenvironment
Renewables & Energy Security
Geology & Landscape NI
Energy & Minerals
Chief Geologist England
Chief Geologist Scotland
Chief Geologist Wales
Geotechnical, Geophysical Properties & Processes
Shallow Geohazards & Risks
Radioactive Waste
Mathematical Modelling
Geological Modelling Systems
Earth & Planetary Observation & Monitoring

4.1.1.3 HYDROGEOLOGY

Environment & Hazards
Groundwater Systems
Groundwater Processes
Groundwater Protection
Urban Geoscience

4.1.1.4 GEOCHEMISTRY

Soils, Landscape & Climate Processes
Geochemical Baselines & Medical Geology
Minerals Sustainability
Mineral Resources & Policy

4.1.1.5 MARINE GEOLOGY

Coastal & Fluvial Environmental Change
Marine Geology
Marine Geohazards & Geotechnology
Continental Margins
Regional Hydrocarbon Prospectivity

Specific responsibilities towards these Teams include:

1. Assessing and approving Data Management Plans and SAN quota requests.
2. Carrying out Team Leader interviews to gather information on ongoing and future data, information management issues and requirements.

4.2 INFORMATION MANAGEMENT IN GEOPHYSICS 2010-11 AND BEYOND (IM1)

Richard Luckett

4.2.1 National and Global Seismic database

The national seismic database contains continuous seismic waveform data for the approximately 100 stations in the British national seismic network, which is operated by the BGS. The constant flow of data into this database from instruments around the country is managed by a series of software packages, which require frequent development and maintenance by BGS staff as the network is expanded and technology advances. Quality control and fault monitoring is done daily and requires a degree of human interaction to support the many automatic checks. The database also contains parametric data for national and global seismic events going back 1000 years. This is the part of the database that underpins BGS's ability to rapidly respond to customers and the media following seismic activity in the UK and abroad. It also has great commercial value, as any seismic hazard work carried out for the UK depends on it. The location and analysis of events before they are included in the database is a large part of work done by the BGS seismology team. The seismic database is part of the National Geoscience Data Centre and of the UK Environmental Observation Framework. The IM1 project supports the maintenance and augmentation of this database and the monthly and annual publication of reports to the customer group - which includes power, water and infrastructure companies as well as local and national government departments.

4.2.2 National and Global Geomagnetic database

The geomagnetic database holds geomagnetic data from BGS observatories in the UK and overseas. It is being constantly augmented with continuous time-series data and derived statistical products. As well as the scientific benefit, such as the possibility of using geomagnetic data as an indicator of solar energy in climate change work, this database underpins the Survey's ability to win significant commercial income (in the region of £0.75M/year). IM1 supports the upkeep and day-to-day growth of this database and the publication of monthly and annual reports.

The geomagnetism team operates two high profile international data centres: A World Data Centre for geomagnetism (see www.wdc.bgs.ac.uk); and an INTERMAGNET Geomagnetic Information Node or GIN (see www.intermagnet.org). IM1 supports work to improve the collaboration between these and other data centres (BGS staff have been leading in this area), to resolve issues relating to data quality between the centres, to further improve the real-time dissemination of data to users and to develop new products (for example, new high-resolution geomagnetic data).

4.2.3 Seismology and Geomagnetism web services

The Seismology and Geomagnetism teams operate web servers in Edinburgh separate from the corporate systems in Keyworth. These are used to rapidly disseminate new data to academics, commercial customers and the public. After an earthquake of public interest these servers can experience many tens of thousands of hits in a matter of hours. IM1 supports keeping these pages secure and up to date as well as work to increase the reliability and resilience of the service.

4.2.4 Recall

The Recall system has been established as the corporate resource for storage of geophysical log data. It is both a database and a set of log display and interpretation applications. Using Recall requires entry of full metadata for every well and log curve loaded into it. This allows for proper tracking of intellectual property rights and data ownership, as well as a better understanding of acquisition detail and log units. IM1 supports the maintenance of Recall and the loading and verification of new data.

4.2.5 Geophysical databases

An ongoing task for IM1 is the maintenance of legacy corporate databases which underpin work done throughout BGS and are key assets of the National Geoscience Data Centre. These include the Stratigraphic Surfaces Database, the Seismic Locations and Sections database and the Seismic Line Index database. Work supported includes the addition of new data and the improvement of user access to these databases.

4.2.6 Impact

It is not possible to cut any of the core project deliverables without significant impact on BGS' ability to deliver on its science and commissioned research objectives. The need for sophisticated data management in these projects has not decreased, if anything it has increased.

4.2.7 Subsidiary deliverables under IM1

4.2.7.1 GEOMAGNETISM PAPER ARCHIVE

There has been a recent increase in demand for historic geomagnetic data (BGS holds continuous time-series data starting in 1848). BGS is not always able to meet this demand because the data has not been digitised (in some cases there is no security copy of the original paper data). IM1 supports an ongoing programme to digitise the whole geomagnetism paper archive.

4.2.7.2 COAL AUTHORITY METADATA

The Coal Authority has deposited large amounts of important data with BGS. This data is in poor condition, on media that are difficult to read, with little or no metadata. Considerable progress has been made to recover this data and this work has now reached the stage where only the more

difficult and time consuming data remain unprocessed. IM1 supports the generation of metadata and its input into Recall for this remaining Coal Authority data on an 'as is required' basis.

4.3 INFORMATION MANAGEMENT IN GEOLOGY 2010-11 AND BEYOND (IM2)

Tim McCormick

4.3.1 SOBI & Borehole Geology

The Single Onshore Borehole Index (SOBI) and Borehole Geology comprise the main source of knowledge of geology at depth used in mapping and 3D modelling by BGS.

FY 2010-11: Work has focussed on the never-ending task of data validation and improvement, promoting coding standards and assigning ‘content codes’ to projects. A trial has also begun of a mechanism to fix the issue of the many-to-many relationship between borehole records (recorded in SOBI) and the actual boreholes.

FY 2011-12: IM2 will continue data validation and providing assistance to geologists using the system. The trial on the SOBI records mechanism will be assessed with the possibility of an implementation of a (moderate) redesign to distinguish boreholes and borehole records.

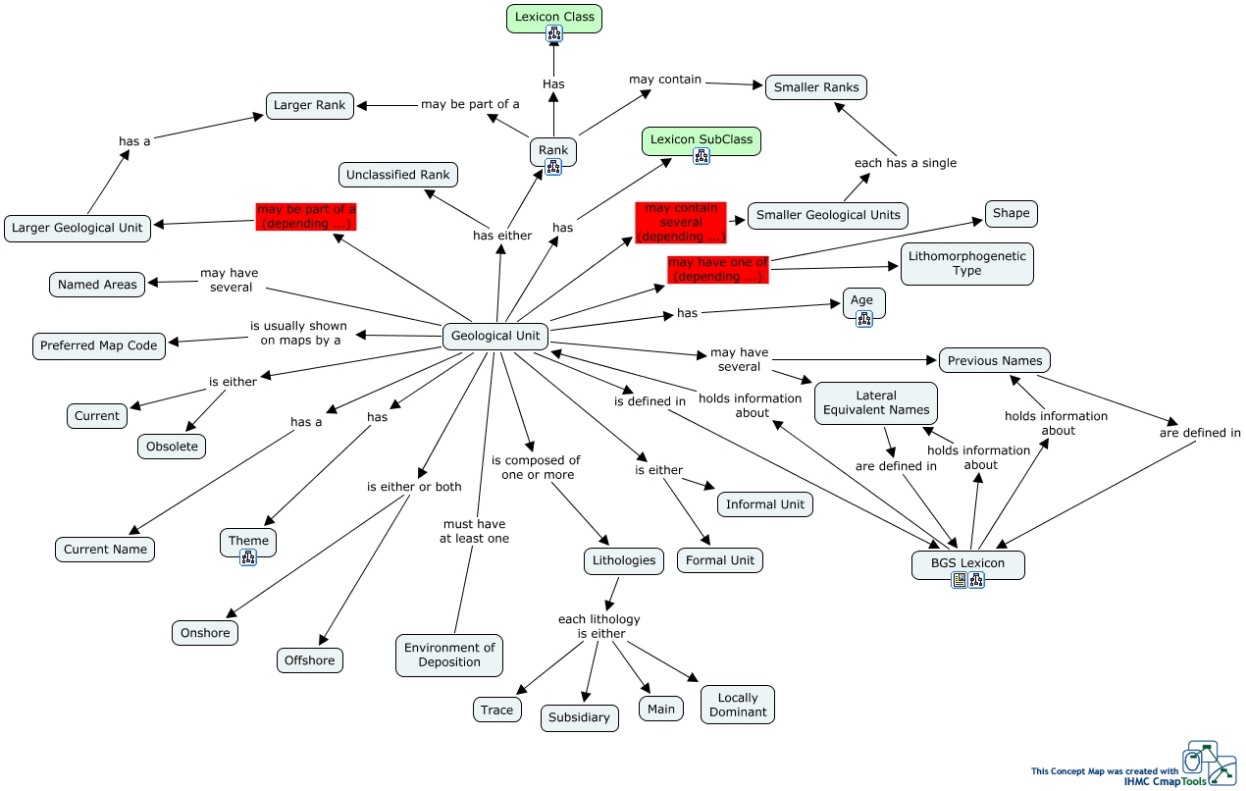
4.3.2 Lexicon of Named Rock Units

As well as being a primary repository of BGS’ knowledge of the UK and its continental shelf, the Lexicon of Named Rock Units underpins numerous corporate workflows and initiatives, including SIGMA field data capture and digital map production, coding down-hole borehole information and 3D geological modelling. It is also available for public search on the BGS website where it is regularly in the top 10 most viewed pages.

FY 2010-11: Work continues to improve the content and coverage of the Lexicon. A pilot test will be made generating short summary versions of definitions, suitable for a non-specialist audience, for use on the BGS website and iPhone app. The Lexicon will be migrated to a revised database (already designed) offering greater information content and functionality. The ‘Sandbox’ system for geologists to author / update Lexicon definitions will be adapted to work with the new database.

FY 2011-12: A major effort is likely to be required to implement the recommendations of the revised BGS classification of lithodemic units on the UK Continental Shelf (UKCS). Work will also be required by Lexicon curators and managers to check a large number of updates from the forthcoming offshore Quaternary Stratigraphic Framework Report. Subject to results of this year’s pilot test, effort may be made to create non-specialist ‘summaries’ of definitions. IM2 will also pilot linking Lexicon definitions to related images of various types (photographs, sketch maps and other diagrams).

Figure 1 Concept map for a definition of a geological deposit, used in developing the redesigned Lexicon of Named Rock Units database.



4.3.3 BritPits

Also known as the mines and quarries database, this contains approximately 130,000 entries for active and inactive onshore mineral workings, and also includes rail depots and wharves involved in movement of commodities. These data are used to produce the 'Directory of Mines and Quarries', and for market intelligence/analysis and resource planning by internal and external customers, including local councils. The data are made available under licence.

FY 2010-11: Work this year includes: loading of data after validation by Mineral Planning Authorities; updating records based on new information received; verification / validation tasks on the existing data to enable production of the interim version of the 'Directory of Mines and Quarries' for the Department of Communities and Local Government (DCLG); ongoing evaluation of the database design to reduce manual data maintenance.

FY 2011-12: Data verification, loading, update and validation will continue, driven in part by the necessity to prepare data for publication of the next full edition of the 'Directory' due in 2011. There may possibly be some minor redesign of the database in response to the ongoing evaluation.

4.3.4 World and UK Mineral Statistics

These used to be two separate MS Access databases, but these were merged into a single system and migrated to Oracle™ in FY 2009-10. As well as being used by BGS for market intelligence and analysis, they are used to produce regular new editions of 'World Mineral Production', 'European Mineral Statistics' and 'United Kingdom Minerals Yearbook', which provide reliable information to policy makers, regulators, industry and the wider public.

FY 2010-11: This year work includes verification, loading and updating of data ready to produce 'UK Minerals Yearbook 2009' (published June 2010), 'World Mineral Production 2005-2009' (due in February 2011) and 'European Mineral Statistics 2005-2009' (due in March 2011). There has also been fine-tuning of the merged database and associated dictionary tables.

FY 2011-12: Data verification, loading and update will continue, driven in part by the necessity to prepare the data for the next editions of 'UK Minerals Yearbook', 'World Mineral Production' and 'European Mineral Statistics'.

4.3.5 Geotechnical Properties Database

The National Geotechnical Properties Database underpins BGS engineering geology research, and is an important information resource for answering enquiries and servicing the data needs of external customers.

FY 2010-11: Following upgrades in the last FY to bring it into line with corporate standards, work continues populating and validating this database. There is still scope to integrate it a little better with BGS corporate standards, for example in referencing rock units.

FY 2011-12: Work will continue in populating the database. Minor modifications may be made to better align it to corporate standards.

4.3.6 National Building Stones Database

This database allows BGS to answer enquiries and provide advice to customers and partners, often local authorities and heritage agencies, on building stones used in the UK. A single National Building Stones database was constructed in FY 2009-2010, in collaboration with English Heritage, merging separate Oracle™ databases previously used in Keyworth and Murchison House.

FY 2010-11: This year, work continues verifying and loading data collected in the field, and fine-tuning the merged database and its associated dictionary tables.

FY 2011-12: There has been a recent change of management of the BGS' building stones work; plans for next FY's activities have still to be finalised.

4.3.7 Dictionary Management

Standardisation and interoperability between BGS databases is ensured by using 'dictionary' tables and other constrained vocabularies to control the terminologies used. IM2 populates and manages the key geological dictionaries, gathering requirements and promoting dictionary use through the Dictionary Management Group which the Project Leader chairs.

FY 2010-11: This year major upgrades are being made to the corporate chronostratigraphy / geochronology dictionary, and a sedimentary environments / mode of origin dictionary is under development. Other dictionaries are being rationalised where duplication or near-duplication is discovered.

FY 2011-12: There is a requirement for a review of the use of lithology description and coding in BGS databases in light of recent and ongoing revision of the BGS Rock Classification Scheme, and the increased use of the Unconsolidated Deposits Coding Scheme. This may drive a redesign to the lithology code dictionaries, and possibly pilot test of knock-on modifications to other corporate databases. The classification of BGS dictionaries and other controlled vocabularies needs to be completed so that they can be offered on the BGS website in a more easily searchable and useful manner.

4.3.8 Geoscience Imagebase 2 (GI 2)

The BGS database of digital geoscientific images, 'Geoscience Imagebase', was migrated to Bright Interactive Asset Bank software in FY 2009-2010 (now known as 'Geoscience Imagebase 2'), and a publicly-accessible subset launched on the BGS website as 'Geoscenic'.

FY 2010-11: This year work continues on verifying the large backlog of image metadata and loading it to GI 2, preparing policies and procedures for the new system, and helping BGS geoscientists prepare their image collections for upload.

FY 2011-12: Work will continue uploading the backlog of images and metadata and helping projects prepare their image collections for upload. A watching brief will be kept on developments relating to the proposed NERC Image Library.

4.3.9 IUGS-CGI Multilingual Thesaurus of Geoscience (MTG)

The MTG working group was instigated under the auspices of the IUGS-CGI in 2003 to revise and improve their existing multi-lingual vocabulary set. The multi-lingual thesaurus will be made freely available and has the potential to be an essential tool for indexing, searching and interpretation to enable interoperability of geoscience data across language boundaries. Along with other initiatives the MTG will be one of the foundations on which to build sophisticated semantic web tools in the future. IM2 represents BGS on the CGI MTG Working Group.

FY 2010-11: Work this year has involved collaborating on population of the thesaurus, and making technical upgrades to the collaborative web database used by the working group members.

FY 2011-12: It is likely that the MTG Working Group will be merged with the CGI Concept Definitions Working Group in the near future. The implications of this for group membership and activities are not known yet.

4.4 INFORMATION MANAGEMENT IN HYDROGEOLOGY 2010-11 AND BEYOND (IM3)

Andy McKenzie

4.4.1 Planned for 2010-11

4.4.1.1 HYDROGEOLOGICAL RECORDS AND WELLMASTER

BGS hold records for over 100,000 water boreholes, in paper form (the National Well Record Archive) and as a digital database (WellMaster). These records underpin many aspects of UK Hydrogeological science, providing the basic information on aquifer properties, water quality and water resources for research, mapping, product development and enquiries. Data are received from drillers as statutory returns under the Water Resources Act, and compiled from research reports and historical archives.

FY 2010-11: Work this year has involved ongoing accession of Water Well Records, scanning of records, and the resolution of indexing and scan registration issues. While WellMaster data entry is substantively complete, there are areas requiring revision and data conditioning has focused on defining aquifer correctly and checking geological details.

FY 2011-12: Work will continue on ongoing accession of Water Well Records, scanning of records, resolution of indexing and scan registration issues to minimize backlogs of data entry. WellMaster aquifer detail revision and checking geological data should be completed.

A substantial collection of paper archives derived from historical groundwater projects will be rationalized and indexed, (subject to staff availability).

Two major donated collections of groundwater data require indexing and incorporation into BGS collections.

Queries to the WellMaster database from the IDA will be reviewed and revised to improve data accessibility, and promote wider use of WellMaster data.

4.4.1.2 WATER LEVEL DATA

The national Groundwater Level Archive (NGLA) holds over 1,000,000 measurements of water level from an observation network across UK aquifers. Data are collected from environmental regulators (and increasingly from NERC research sites). The archive complements CEH EIDC activity on surface water. Data are regularly collated into monthly and annual summaries, and the collection is extensively used as a research resource. The operation of the network is overseen by the government Surface and Groundwater archives committee.

FY 2010-11: Monthly capture of groundwater level data and reporting has continued. New telemetry equipment is being commissioned on BGS research sites, and data from these sites will be integrated, in real time, with the archive.

FY 2011-12: Monthly capture of groundwater level data and reporting thereupon will continue. Building on analyses carried out by a recently completed research project, gaps in data records will be in filled by modelling, if sufficient data can be identified.

4.4.1.3 HYDROGEOLOGICAL DATA

Other hydrogeological datasets include data on water quality, geophysics and data from specific research projects. Wherever possible these data are incorporated into existing corporate collections, but the development of new areas of research, especially cross-disciplinary research, require new data management approaches.

FY 2010-11: Development of data management for a new collection of groundwater fauna, merging biodiversity, ecological and geological information.

FY 2011-12: Review of data collection and data holdings associated with groundwater and climate change.

Several hydrogeological datasets requiring curation (generally improved indexing and integration with corporate collections) and development will only be addressed if staff resources are available in future years. These include the geophysical, geochemical and aquifer properties datasets, where data are available to end users, but access is predicated on assistance from staff with specialized knowledge of data structure and location.

4.5 INFORMATION MANAGEMENT IN GEOCHEMISTRY 2010-11 AND BEYOND (IM4)

Susan Hobbs

4.5.1 Current 2010-11 tasks

1. Refinement of Query Layers for the geochemistry database to improve ease of use is ongoing. Training and familiarisation of the new structure of the database will be provided.
2. Reformatting of data from the Laboratory Information Management System (LIMS) in preparation for entry into the geochemistry database. Transfer of the chemistry laboratory data from a local sever to the W: drive.
3. Capturing new and legacy data from chemical analyses and maintaining the data quality through continuing audit.

4.5.2 Planned for end of 2010-11

1. Work shop to provide familiarisation with the new structure of the geochemistry database for frequent users.

2. Transfer of data form GSF\LaboratoryOperations to SciFac\AG should be completed by the end of the FY. Outstanding data currently in the LIMS should be loaded to the geochemistry database also by the end of the FY.
3. Auditing and capturing of new and legacy data including 2008 and 2009 water data. Reports documenting background information on the collection, condition and analysis of the samples.

4.5.3 Planned for 2011-12

In addition to the ongoing geochemistry database maintenance and enhancement, IM4 will cover data management in the following areas:

The current LIMS is inflexible and complicated to use. A new LIMS is to be purchased and will require the creation of an interface/loading program to the geochemistry database to replace existing programs.

Continue with auditing and capturing of new and legacy data. The 2010 water samples (up to 2000 sites for approx 60 analytes) collected and analysed in 2010-11, will be audited.

Investigate the inclusion of a sample processing field in the geochemical database. For some methods of analysis, this is as important as the method of analysis. The field would require the creation of a dictionary of sample processing methods. This work is to be done in collaboration with staff from Geochemical Baselines and Database Solutions.

If resources allow, data from Monitor water will be loaded to an Oracle table and linked to the geochemistry database.

Radon data currently sitting in project files will be moved to the digital archive if resources allow. The data is complex and current resources are insufficient to create a database of the necessary complexity.

4.5.4 Geochemical data at the British geological survey – report findings

Geochemistry Report – Geochemistry@BGS: A guide to geochemical data at the British Geological Survey, OPEN REPORT OR/09/028, identified various issues as requiring attention.

The following recommendations were made in the report to address these issues (text in italics is additional background explanation):

1. Project Data Management plans need to be managed and enforced to ensure projects generating geochemical data conform to corporate standards for the storage of samples and data. This should be a function of the Geochemistry Data Management Project but will only be successful if projects are provided with clear guidance and policies as to what they should do with the data that is generated.
2. With the exception of the GeoIndex, links on the BGS internet about geochemical data and information are dispersed and hard to find. There is an absence of any high level entry point providing links to the existing geochemical resources. The removal of the word “geochemistry” from all parts of the organisational structure leaves users without any reference point to start from. A search on “Geochemistry” gives a long list of unjoined-up links that is an unhelpful overview of geochemistry at BGS. It is suggested that links to all the geochemical theme areas in this report are focused from within one specific science area of BGS.
3. It is misleading to have something called the corporate BGS Geochemistry Database that only receives a small part of BGS’s geochemical data. External (and indeed internal users) are misled into believing this is BGS’s only source of geochemical data. Whilst it would be good to have all BGS geochemical data in one place this would be impractical

(see below). It is suggested that the term "corporate BGS Geochemical Database" be qualified to reflect its true content, *e.g.* corporate BGS Geochemical Database for the UK land surface. *The Geochemistry database was not designed to be the "corporate" BGS Geochemical Database. It was designed initially to be the Geochemistry database for the Minerals and Geochemical Surveys Division. As the Divisions disappeared and boundaries merged this distinction was lost and an assumption was made that it was the Geochemical Database for the whole of BGS. The name should be changed to indicate the actual coverage.*

4. Ideally all BGS geochemical data should be accessed via linked Oracle databases. However, this would require the investment of a lot of resources to replace and revamp adequate databases that already exist in different areas across BGS. The more pragmatic approach is to focus on delivery of data to the users. This will mean the creation of derived geochemical products delivered via the internet. An example of this would be a set of ArcGIS geochemical images for UK baseline geochemical data downloadable from the BGS web site. *At the moment the GeoIndex only provides the location of the sites along with a list of analytes determined.*

5. The licensing of land surface geochemical data for the UK has been operating satisfactorily for many years. There needs to be a consistent pricing policy for all geochemical data, particularly with regard to the provision of geochemical data for external academic research. *Any changes will need to fit in with the new NERC data strategy.*

6. Some particular areas that have generated geochemical data need support to capture legacy data and information before it is lost. Currently much of the data from estuarine, radiometric and raw G-BASE geochemical data is dependent upon the custodianship of individuals rather than corporate systems. The capture of these data sets to corporate systems needs to be addressed in the information management delivery plan. *The estuarine data is the remit of IM6 (Marine). It was part of this year's delivery plan to investigate the radiometric data, however due to cuts at the beginning of the year there were no longer the funds to proceed with this. Since 2004 raw G-BASE data has been loaded to the "Raw" data tables in the Geochemistry database. Will need to look into loading raw data prior to 2004.*

7. Currently estuary geochemical data is neither "owned" by the Marine Geology or Environmental Geoscience Baselines themes and as a consequence, data from estuarine projects is in danger of becoming neglected. This report defines offshore as being below the high water mark and consequently estuary data is considered as being "offshore".

8. This report has demonstrated that in many areas of geochemical data generation it is publications and reports that play an important part in information dissemination. This is particularly the case for isotopic, organic and international research projects. The hydrogeology portal gives excellent access to downloadable pdfs of many hydrogeochemistry reports but this is restricted to BGS intranet users only. Bibliographic lists like those provided in the hydrogeology portal should be made available from the BGS internet site with associated links to free downloadable reports. There is a great wealth of geochemical information and data in inaccessible, non-digital internal and technical BGS reports, particularly for international research projects. These should be scanned and made available over the intranet. *The entire series of Mineral Reconnaissance Programme (MRP) reports are now free to download on the MineralsUK website (pdf format). In total there are 145 reports and 16 data releases.*

9. EndNote bibliographies provide useful and comprehensive lists of reports that detail geochemical information and data. A corporate approach to the use and sharing of EndNote bibliographies is required. Valuable EndNote libraries are lost when programmes/Science theme areas are reorganised.

10. The confidential classification on many BGS technical and internal geochemical reports needs to be reviewed and wherever possible removed. Many of the geochemical theme areas are already making good use of the NERC Online Research Archive (NORA) but more geochemical reports should be added.

11. Team Leaders associated with the various geochemical theme areas should be providing more "geochemical data awareness" lectures within BGS and externally (for example to university departments). There is also scope for the provision of a training course to educate non-specialist geochemists in the use and interpretation of geochemical data. *A work shop to aid staff with the extraction of data from the "Geochemistry database", using Microsoft Access was run in December and also in the previous year. This workshop also included using ArcGIS and Minitab.*

12. Legacy Field data is to be added when and where funding allow.

4.6 COASTAL AND MARINE INFORMATION MANAGEMENT (IM6) 2010-11 AND BEYOND

Paul Henni

4.6.1 Principle Objectives:

- Management of the marine geoscience data system and team, integrated with the NGDC, capable of unifying the disparate data types and formats for long-term storage, preservation and delivery, with appropriate documentation.
- The underpinning of BGS marine geoscience, including ongoing data acquisition and the management of data and information for internal and external use.
- The continuing improvement of access to data and integration of marine geoscience data with BGS corporate data management systems and international standards.
- Development of methods for capturing new and innovative data types, developing methods for data for new and emerging requirements.

4.6.2 Completion and maintenance of a fully integrated Marine Data Management System to corporate and international standards

4.6.2.1 BGS Marine Data Management Standards for Offshore Surveys

The BGS Marine Data Management (DM) Handbook is available as an Open Report and provides overall guidance. Internal data management, QC and validation procedures have been developed and documented from existing TWIKI collaborative work and are now active.

4.6.2.2 Accessioning (incoming data and information)

We have worked closely with NGRC to ensure that accessioning (booking-in) of new marine data takes place. Crucially, this process includes details of terms and conditions of access and use of these data; removing the reliance on individual knowledge in favour of an appropriate level of open documentation.

4.6.2.3 The Coastal and Marine Database (CMD)

This is a spatially enabled data repository linked to BGS online SAN storage, the BGS offline Digital Archive and to the delivery of discovery and evaluation metadata. Work has been ongoing to upgrade the design, including technical metadata, together with enhancements to add to the data types dealt with and to allow the migration into the CMD of legacy sample and report

data. Under this work, IM6 are also working to improve corporate dictionaries, such as Lithostratigraphy, to ensure onshore:offshore compatibility of data to BGS standards.

4.6.2.4 CREATION OF METADATA TO MEET REQUIRED STANDARDS (INSPIRE, MEDIN, GEO-SEAS)

IM6 have been responsible for the publication of metadata to the Marine Environmental Data and Information Network (MEDIN) standard via the NERC Data Grid. Additionally, test data has been created to deliver to the Geo-Seas initiative. This work will continue, together with linked web data delivery, in 2011-2012.

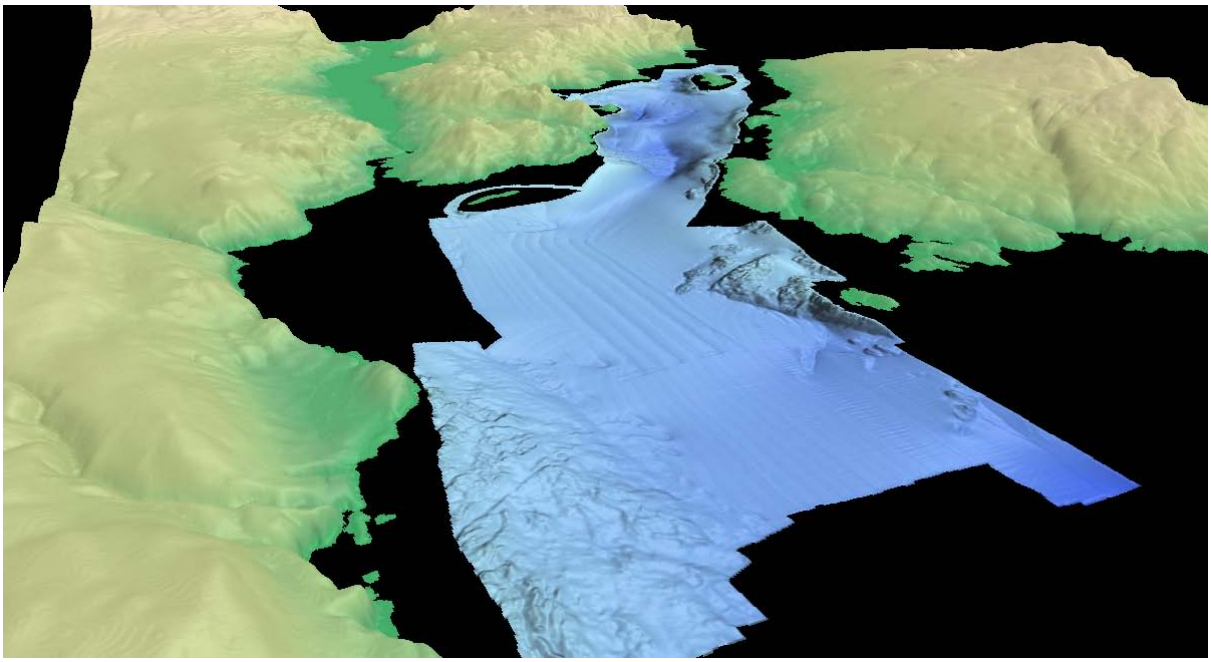
4.6.2.5 DELIVERY OF DATA AND INFORMATION

A draft document is in preparation identifying what data meet NERC's Data Policy and should be delivered for free (or at basic handling cost), ideally, where practical, via the web using the BGS Offshore Geoindex. IM6 is closely involved in the creation of the required indexes, files and delivery mechanisms, in consultation with the BGS Information Delivery Programme.

4.6.2.6 DATA MANAGEMENT EXAMPLE

BGS collect or receive Multibeam 'Digital Terrain data' showing the seabed for geological interpretation. These datasets can be several hundred Gigabytes in size and must be stored appropriately for re-use.

Figure 2 Image shows BGS data collected in Loch Eribol, north Scotland, showing (blue) offshore multibeam and (brown/green) onshore terrain, looking approximately southwards. (Image created by R. Cooper, BGS)



4.6.3 Other areas of work under IM6

4.6.3.1 Continue to operate as a MEDIN Data Archive Centre (DAC) and maintain active participation in MEDIN (including in-kind BGS contribution from IM6 funds)

IM6 are responsible for bidding for MEDIN funds and for co-funding this activity to achieve the long-term archiving and delivery of marine geology and geophysics datasets of national importance.

4.6.3.2 Ongoing processing of new and immediate legacy (post-2000) BGS or combined BGS/Partners marine surveys

It is planned that all post-2000 BGS and partners surveys will be signed off as completely data managed by the end of FY 2011-2012.

4.6.3.3 Data Rescue

Examples of this include the scanning and, where appropriate, vectorisation to digital SEG-Y of legacy paper Seismic records. This is a very wide area of potential work, so is being led by user community requirements, with priority given to those data highlighted by BGS Team Leaders.

4.6.3.4 United Kingdom Hydrographic Office (UKHO) data and information 2011-12

This is a major dataset of existing BGS-held legacy and incoming new data in both analogue and digital format. A MoU (Memorandum of Understanding) has been signed between the BGS and UKHO and IM6 is working closely with NGRC to implement a plan for both review and partial disposal of old analogue records, together with the receipt of appropriate and useful 'new' records and digital data.

4.6.4 Other ongoing work planned for 2011-2012:

- Develop further the Marine Enquiry procedures to facilitate transfer of routine enquiries across to the BGS Central Enquiries Team
- Continue the passing of the management of analogue data and paper records to NGRC colleagues, providing advice and knowledge transfer
- Continued development of the Marine Data Management Team and subsequent knowledge sharing
- Work with the Landmark Manager to improve knowledge and data management of this system and data, including legacy tape issues
- Complete definitive set of BGS-held BIRPS data and information (subject to staff time for and availability of required specialist)
- Publicise BGS data and information holdings

4.6.5 Desired Tasks 2011-2012, subject to funding

- Complete review and integration of the disparate 'Marine' and 'Coastal' analogue and digital collections into a unified and coherent indexed and delivered set of information (working across all BGS areas of work).
- Legacy projects and material received from retired colleagues.
- Legacy survey data and information pre-2000
- Non-BGS survey data and information (outwith MEDIN DAC data)

5 European Co-Funded Activities

5.1 GEO-SEAS

Helen Glaves

5.1.1 Overall objectives

1. Creation of unified European e-infrastructure for marine geoscientific data
2. Facilitate the locating, accessing and delivery of federated marine geological and geophysical data sets
3. Provide user access to harmonised marine geological and geophysical metadata and data via the Geo-Seas portal
4. Develop interoperability with other data types within other disciplines

5.1.2 Current tasks 2010-11

5.1.2.1 INSTALLATION

All data centres are currently installing the Geo-Seas software components including MIKADO, a dedicated Common Data Index (CDI) metadata creation tool which is freely available and widely used in the oceanographic data community; NEMO for converting datasets to the agreed delivery and exchange formats; and the Download Manager which along with the associated Request Status Manager (RSM), connects each of the 26 data centres to the Geo-Seas portal and allows users to query the CDI metadata records and directly access the datasets held by each of the data centres. All data centres are required to install this set of standard tools in order to become a fully functioning Geo-Seas data centre.

The installation of the Geo-Seas tools at each data centre is a phased process which started in August 2010 and will be completed for all data centres by the end of May 2011. The initial group of data centres included those that were already experienced in the use of the Geo-Seas tools through participation in the related SeaDataNet project. The second group is comprised of those users who have experience of using similar software tools and can quickly undertake the installation of the Geo-Seas tools. The final group are those data centres who are inexperienced with this type of tool and who are likely to require additional support during the installation process.

In addition to this basic suite of tools a number of the Geo-Seas data centres are installing additional software components to allow these partners to provide extra advanced services. These additional services include 2D mapping services and also 3D viewing and analysis services.

5.1.2.2 POPULATION

The Geo-Seas partners are currently engaged in creating CDI metadata records, an enhanced form of the ISO19115 metadata standard, for each of the datasets which they hold. These CDI records are then being validated and loaded to the Geo-Seas database. All data centres have completed the initial phase of creating a trial set of metadata records which have been used to test the process of metadata creation and data delivery by the individual data centres.

Those Geo-Seas data centres who are already familiar with the Geo-Seas tools through experience with other projects have provided full metadata records and the associated data sets

which are now accessible via the Geo-Seas portal. The fully functional Geo-Seas portal is available via the Geo-Seas website http://www.geo-seas.eu

5.1.2.3 DISSEMINATION

Geo-Seas partners are actively promoting the Geo-Seas project through presentations at various international conferences including the European Geological Union Congress, American Geophysical Union and International Marine Data and Information Systems conference.

5.1.3 Deliverables

Geo-Seas is working to a schedule of deliverables as defined in the Description of Work (DoW). The deliverables for the period 2010 – 2011 are shown in Table 5 below.

Table 5 Geo-Seas Deliverables 2010 - 2011

Del. no.	Deliverable name
1.2	6-monthly progress reports
1.3	Minutes of Steering Committee meetings
1.4	Annual progress reports, including cost statements
2.1	Minutes and actions of Full Network meetings
2.2	Infrastructure Monitoring Plan with indicators and measurement protocols
2.3	Operational Monitoring system
3.1	Training course 1 for data managers: programme and curriculum
3.2	Report on the training course 1 with evaluation
3.3	Training course 2 for technicians: programme and curriculum
3.4	Report on the training course 2 with evaluation
5.2	Initial production of CDI records integrated in the on-line system
5.6	Contributions from Geo-Seas data centres in EDMO, CSR and EDMERP
5.7	Analysis and plan for upgrading of SEISCANEX metadata into the CDI directory
8.1	Adapted and tested Geo-Seas components for entry, production and exchange
8.2	Adapted CDI User interface and transaction system for Geo-Seas portal
9.1	Implementation plan for the gradual installation in phases
9.2	Set of instructions and documentation for basic services installation
9.3	Pilot group of Geo-Seas centres operational with basic services
9.4	Evaluation and revised set of instructions and documentation for basic services installation
10.2	Prototype software for seabed and geological log viewing
10.3	User feedback and required adjustments
11.1	Prototype software for high-resolution geophysical viewing services
11.2	Test report on prototype software

Del. no.	Deliverable name
11.5	DTM Prototype delivery for evaluation by users
11.6	DTM User feedback and required adjustments

5.1.4 1st Annual Review

The first Periodic Review meeting between the Coordinators, Wim Jansen (EU Project Officer) and an independent panel of reviewers, selected by the European Commission, took place on 21 June 2010. The purpose of this review meeting was to assess the progress which has been made during the first year of the project. The outcome of this review was successful with the panel being satisfied that the project had completed all deliverables, which were scheduled for this reporting period, to a good standard. The project had also achieved all the stated milestones as defined in the project Description of Work.

5.1.5 Geo-Seas training course

A data manager and a technician from each of the Geo-Seas data centres attended a training session which was held on 22 -25 June 2010 at the University of Strasbourg, France. The purpose of this training was to provide instruction on the installation and use of the Geo-Seas software tools and the data management standards being implemented for the participating data centres.

5.1.6 Planned by end of 2010 -2011

All data centres will have installed the Geo-Seas software tools by the end of February 2011 and will have begun a full-scale population activity which includes generation of Geo-Seas metadata records and making the associated datasets available in the agreed delivery formats. As a result all data centres will be fully connected to the Geo-Seas portal By the end of July 2011. In addition the associated vocabularies will also have been updated with the additional terms which have been identified as necessary as a result of this population process.

Geo-Seas is currently exploring opportunities for collaboration with other projects and organisations beyond Europe. Following the receipt of a Foreign and Commonwealth grant which allowed a 10 day visit to Scripps Institution of Oceanography, San Diego, USA (SIO) work is now underway to develop a 'proof of concept' for the delivery of metadata from the R2R project, a joint initiative between SIO, Woods Hole Oceanographic Institute, Lamont Doherty Earth Observatory and Florida State University, via the Geo-Seas portal. This proof of concept will be used to develop a proposal for US funding of this activity allowing R2R to participate in the Geo-Seas project.

A splinter group meeting is planned for the European Geological Union meeting in Vienna on 7th April 2011 The meeting will be attended by representatives of a number of relevant EU projects and colleagues from the US and Australia. This will be an information session with a view to developing European:US collaborative efforts and proposals for future funding.

In addition links between Geo-Seas and the European Marine Data and Observation Network (EMODNET) Geology will also be developed which will include integration of the Geo-Seas data with the data products being developed by the EMODNET Geology.

5.1.7 Planned for 2011-2012

Additional functionality will be provided for all the Geo-Seas data centres via the portal with geophysical and geological log viewing services being made available.

Geo-Seas funding will end in October 2012 at which time all deliverables and milestones outlined in the Description of Work will be completed. In addition the project consortium will develop a proposal for follow-on funding as part of the EU call for proposals which is scheduled for 2012. The Coordinators of the project will be negotiating with the European Commission in advance of this call to ensure that it is orientated towards the refunding of the Geo-Seas project.

A final 'wrap-up' meeting for the project will take place during September 2012 in Cork, Ireland. This meeting will take the form of an international workshop held over three days and which will include delivery of the project results. An initial announcement of this meeting will be made in September 2011.

6 Additional Tasks

6.1 STRATEGIC DELIVERABLES

In the financial year 2010-11 unallocated staff could be bid for on "strategic deliverables". Each Team Leader being provided with a project code against which bids could be made. The bids were authorised by the relevant Science or Information Director. During this year the following tasks were undertaken as strategic deliverables under IMFO:

6.1.1 Marine Data

There have been significant amounts of data generated by a large number of Marine Surveys and there is still considerable work to be done to ensure the quality of these. This Survey work is being broken down into achievable work packages to ensure steady, measurable progress. The longer the data is left without thorough quality control checks being carried out, the more time will be required to retrospectively do the work, so it is important to maintain the current momentum.

The objective is to provide enhanced, quality controlled Marine Geoscience data, stored to corporate standards, with appropriate elements stored in the BGS offline archive system. This will lead to improved delivery of revised datasets to internal and external customers.

Specific deliverables:

1. Successful delivery of MEDIN-compliant metadata to the MEDIN web portal and updated BGS Offshore Geoindex.
2. Completed review and update of post-2000 Marine Surveys. Initiation of work on legacy surveys (pre-2000).
3. Up to date and quality controlled Marine Sample information (clear 5 year backlog of data loading).
4. Support of the Marine Mapping Programme, completion of scanning, archiving and vectorisation to SEG-Y of seismic records (for appropriate data types) of target offshore regions: Forth Approaches, plus others to be agreed with Marine Team Leaders.

6.1.2 Data Recovery and Conditioning

"In the standard model, one collects data, publishes a paper or papers and then gradually loses the original dataset."

- Geoffrey Bowker

(Gibson and Lord 2008)

The aim of this task is to ensure that project information assets currently lying inactive and unused on the SAN are recovered and made ready for use and re-use and incorporated into corporate records where they form the basis of products and services. In short, recovering assets from completed projects and conditioning them for further integrated use.

The OldProjects area of the SAN holds ~ 200 closed projects, with a combined budget allocation of ~ £10,000,000, dormant and awaiting asset recovery. They occupy over 2.5 Terabytes of SAN storage. Two years ago there were 95 projects occupying 1.5 Terabytes. There are further closed projects within the Team structure.

Information assets from these projects must be assessed, conditioned and incorporated into corporate databases, products and services.

The digital archive and data management procedures have improved since two years ago but staff need to be guided in using the systems.

Research data are a valuable resource, often requiring much time and money to be produced. Many datasets have a significant value beyond the original research. These outputs should not be "gathering dust" but should be actively managed. Data must meet quality thresholds and be stored in a reliable manner supporting their promotion and reuse.

Well promoted quality data increases the impact and visibility of research, promotes the research that created the data and leads to new collaborations. At the same time, BGS needs to ensure that expensive data storage systems are efficiently managed and not used to amass abandoned data.

Examples of specific tasks undertaken include verification activities on SOBI and associated borehole images, and recovery of specific data holdings such as groundwater models:

BGS has been developing groundwater models for a number of decades. In the last ten years, the groundwater programme has significantly increased its output of groundwater models. Recently, there has also been an increase in staff members working on groundwater modelling projects. Therefore, to ensure that duplication in effort did not occur, it was deemed necessary to create an inventory of groundwater models. This has taken the form of a shapefile consisting of outlines of the model boundaries, if readily available, or rectangles showing the full spatial extent if a digital boundary file did not exist. A metadata catalogue has also been created that relates to basic information on the model, its purpose, how it was funded as well as where further information about the model can be obtained. The aim for next financial year is to setup an archive on the S: drive for all groundwater models created within BGS as well as making the shapefile of model outlines available externally via the BGS website.

6.1.3 Environmental Observation Framework

This task was to update the records within the Environmental Observation Activity Catalogue on behalf of the UK Environmental Observation Framework (EOF). The task was under the direction of Andrew Howard, the BGS contact for the EOF.

It is intended that this catalogue will hold information on all environmental observation programmes and activities carried out by or for the UK (i.e. those covered by the public, private and voluntary sectors). It will contain information about current and future activities, and past activities where outputs are still available for use.

UK-EOF definitions

Environmental: *the broadest sense of observations from the natural environment concerning physical (including geological), chemical and biological properties of the environment. This includes observations collected on land, in air, in ice, in freshwater and in the coastal and marine environment, compliance or statutory information, Earth observations from space and the effects of humans on the environment. Note the exception is human health data.*

Observations: *the taking, on a reasonably regular basis, of any form of observations relative to the status of the environment, regardless of frequency of, or purpose for which, the observations are made, or however they are made (from satellites, ships, etc). Such observations are designed to meet a wide range of societal needs by providing a variety of products and services. Surveys are in scope for some work streams.*

Programmes: *a programme is a collection of projects or elements of observations; programmes may consist of several activities*

Activities: *an activity is a single project or element*

Relevant information on BGS environmental monitoring activities was collated and reported to the EOF in the format requested.

6.1.4 Geoscience Imagebase (Asset Bank software)

The successor to Geoscience Imagebase was purchased, customised and installed during 2009-10, and is now jointly managed by Publications Services and Information Management. Additional recent customisations have been delivered by the supplier, Bright Interactive. Detailed implementation work is required before the full system can be rolled out internally. The work involves:

1. implementing the management functions; setting up the extensive user customisable options; designing and documenting workflows; rolling the new version out to the managers of the system and geoscientists;

2. writing customised help for the system;

3. coordinating and managing clearing the backlog of images and metadata that have accrued while the system has been offline.

Specific tasks completed have included:

a. ~ 16,000 images/metadata sorted and uploaded to Geoscience Imagebase. Twelve special collections have been processed through to the Public 'Geoscenic'.

b. Work with the Landslides Team to sort out their project area images. Batch 01 (~ 2000 images has been fully processed (images and metadata will be pushed to Geoscenic imminently). Batch 02 ~ 4,600 images - images extracted, registered/renamed/uploaded to the Geoscience Imagebase. Currently working through metadata.

c. The Pre-Migration to Assetbank. Master tiff images have been converted to jp2 and the Tiffs sent to Tape Archive. Work on the delivery of the full resolution image via jp2 viewer, integrated into the Geoscience Imagebase, is underway.

6.1.5 Lexicon Rollout

The Lexicon of Named Rock Units is one of the most popular resources on the BGS website, represented in OpenGeoscience and regularly in the top 10 most viewed pages. However it has a dated 1990s appearance and is lacking in functionality.

The redesigned Lexicon database stores additional attributes and presents them in a more useful way. This work will complete the upgrade of the Lexicon search screens on the IDA, including the geologist's 'Sandbox' system for authoring and updating definitions. The search screens will be duplicated on the BGS website, greatly enhancing its usefulness and service to customers.

Specific deliverables:

1. Rollout of redesigned Lexicon of Named Rock Units database and migration of data.
2. Completed Lexicon search screens for the new database on the IDA.
3. Completed geologists 'Sandbox' for authoring and updating definitions in the new Lexicon.
4. Database and search / display screens duplicated to the BGS website with robust mechanism for regular refresh of the external data. Include display of summaries of geological units suitable for 'popup' from GeoIndex and Geology of Britain Viewer.

6.1.6 Archiving CA Seismic and Well-log Tape Data from the SAN

In previous years Coal Authority seismic data were received under a task in IM1. These data were stored as received in project space on the SAN. The organisation of the data in this way was not optimal for re-use or the answering of enquiries. A proposal was put forward to reorganise the data in a way which would provide more user-friendly and efficient access long-term. Appendix 1 contains the detail.

6.1.7 Lexicon Summaries

This task covers a pilot study to create summaries, suitable for the lay-reader, of selected British rock units for display on the BGS Website, Geology of Britain Viewer, iPhone and similar applications, and which can form the basis of GeoReports descriptions. These texts will be stored within the Lexicon of Named Rock Units and based on information from that database and other sources. Units appearing on the BGS UK Stratigraphy Charts will be prioritised, and Lexicon entries will be synchronised with the charts. The work will coordinate with Web Systems and Enquiries to ensure outputs suitable for web and GeoReports purposes.

The output will be concise summary descriptions of key rock units and deposits on the UKCS, suitable for a non-specialist audience, which can be displayed by the BGS website and iPhone app, and which can form the basis of GeoReports descriptions, thus removing the need for Geo-reporters to repeatedly recreate their descriptions. As a by-product, entries in the Lexicon will be checked against the BGS UK Stratigraphy Charts and improved.

Specific deliverables:

1. Summary descriptions of UKCS rock units and other deposits. The Lower Cretaceous, Lower Jurassic and Mercia Mudstone have been suggested as the focus for a pilot study. Information sources include the Lexicon, Stratigraphic Framework Committee (SFC) reports and GeoReports.

2. Texts that can form the basis of GeoReports descriptions for each rock unit. This may be the same text as described in 1 above, or a separate but closely similar piece. This will be determined as the pilot study progresses.
3. Upgrade of Lexicon definitions where this can be done as a 'quick win' using the same information sources. As a minimum, age, lithology, hierarchical relationships, and environment of deposition / mode of origin are to be filled in.

7 Conclusion

The IMFO Team is engaged on a wide range of data activities which underpin the scientific research and commercial activities of the BGS. The work is essential for the long-term security and validity of the most fundamental and basic data central to the effective working and reputation of the organisation.

Results of research activities must be managed effectively to support current and future research and to assure their quality, accessibility and security. Without ongoing data management to agreed corporate standards, crucial information that underpins current and anticipated BGS research, product development and advice provision will remain incomplete and may become lost, devalued or corrupted.

Specialist data management across the five science themes provides a unifying element to the myriad project data output streams; building on single elements to produce a more integrated resource greater than its individual components. The corporate verification and metadata functions improve data quality and access. Data management planning raises awareness of important data issues when planning and running a project and ensures they are considered.

Data produced as a result of the thousands of pounds spent on research should provide the foundation for future research; they need to be seen but, crucially they need to be fit to be seen. Data outputs should be regarded as at least as important as the research reports themselves if the research community are to build on the successes of the past.

Appendix 1 Archiving CA seismic and well-log tape data from the SAN

John Rowley & Christopher Royles, 5 January 2011

PREAMBLE

Data recovered from Coal Authority seismic and well (borehole) tapes is stored in the SAN directory:

W:\OldProjects\CASeismicSurvey\

Although subsets of this data are accessed regularly, mostly to satisfy external data enquiries, usage of the data set as a whole is rare. We therefore propose to re-organise this data set into data packages which can be exported to archive storage and then retrieved as needed for data enquiries.

The most convenient way of packaging the seismic data is on a coal prospect (exploration and production area) basis. This is the way the data has always been managed by the NCB, BCL and CA; and is almost always the way it is requested by enquirers.

Well data was organised by well name by the NCB, BCL and CA because borehole drilling operations were not as closely tied to coal prospects as seismic surveys, with many boreholes drilled for colliery development. External data enquiries typically list well names. We therefore propose to package well data by well name.

CURRENT DATA ORGANISATION

Seismic data was recovered from original open-reel and cartridge tapes in two main recovery projects, both funded directly by the Coal Authority:

- In **1999** all stack data and demultiplexed field data tapes were identified and recovered to archive-quality DLT tape cartridges. This data was copied to the SAN in 2007, because problems had been encountered when reading some of the cartridges, and because data retrieval time from the cartridges was identified as causing significant cost and delay in delivering data to customers. This data is organised: one directory per DLT cartridge tape, one file per original tape file.
- In **2008** all remaining tapes were recovered to portable hard disk storage. The majority of the data was from seismic field recordings. Two outputs were produced from each tape: (1) a copy of the data in its original instrument recording format; (2) a demultiplexed copy of the data in industry-standard SEG-Y format. The latter is that most commonly required by data customers. This data is organised: one directory per shipping pallet, one sub-directory per shipping box, one file per original tape.

In addition, there is some seismic data from individual customer-driven tape data recovery projects which took place between 1999 and 2008. Data was returned to BGS for the CA archive on various media (Exabyte tape cartridge, CD-R and DVD-R). In some cases this work included scanning of acquisition documents (e.g. observers' logs, surveyors' lists and omission summaries).

Well data has been recovered from original tapes on an ad hoc basis, as funding allowed:

- In **1998-9** about one third of the well tapes were read at BGS Keyworth. Data was output in the 'pre-load' format used by our old in-house WELLOG database. All the files were

archived in the Tivoli storage management system (TSM), with data for key wells being loaded in WELLOG.

- In **1999 – 2004**, a few well tapes were read on a customer-driven basis. Data was output in original formats, and typically returned to BGS for the CA archive on CD-R media. Most data was loaded into WELLOG.
- In **2004** data was recovered from all remaining unread well tapes by a subcontractor in a BGS-funded project. Data was output in original formats on CD-R media. Recovery from older media was patchy. Most recovered data was loaded into WELLOG.
- In **2008** all well data tapes found in the CA archive were recovered to portable hard disk storage in the same CA-funded project as the seismic tapes. This included about 400 tapes which had not been in the CA tape catalogue. Data output was in original format and organised: one directory per shipping pallet, one sub-directory per shipping box, one file per original tape. This duplicated earlier work, however data recovery was generally found to be better.

NB. WELLOG has been replaced by a commercial well log database package, Recall, and all data transferred.

PROPOSED DATA REORGANISATION AND ARCHIVING

Seismic data is usually accessed by prospect (exploration area), for external enquiries and internal BGS use. Therefore we propose to reorganise the data on a per-prospect basis, gathering together:

- Stack data, from the 1999 recovery project
- Demultiplexed field data, from both 1999 and 2008 recovery projects
- Raw field data, from the 2008 recovery project
- Any other data (scanned obs logs, etc.)

The resulting data structure would be:

- Directory: named for the prospect
- Subdirectories for each data category

Well data is usually accessed by well name, for external enquiries and internal BGS use. Therefore we propose to reorganise the data on a subdirectory-per-well basis, including:

- Data from the 2008 and earlier data recovery projects
- Any scanned and digitised geophysical logs (these are created when customers require data but tapes are unreadable, were missing or were never produced)

The index spreadsheets for the recovered seismic and well data sets contain all the relevant information required to construct the directory structure and allocate individual data files to the correct subdirectories within it. The spreadsheets can be sorted and filtered to drive scripts which rename and move the data from the current structure into the new structure.

Statistics

Coal prospects:	190
Stack tapes:	1,100
Demultiplexed field tapes:	2,000
Raw field tapes:	10,110
Well data tapes:	2,150
Approx. data volume (GB):	800

Workflow

1. Cross-reference indexes to ensure a common set of prospect and well names (same spellings. etc.)
2. Index scanned seismic acquisition and well data (observers' logs, etc.)
3. Seismic: 2008 Data and 1999 Data
   - 3.1. Prepare scripts to rename and move data files into archive structure
   - 3.2. Sort and filter index spreadsheets to build data lists for each prospect and data type
   - 3.3. Run scripts with each prospect data list; check processing logs and manually move any problem files
   - 3.4. Archive prospects as completed.
4. Well logs:
   - 4.1. Cross-reference indexes from various copy jobs (multiple copies become versions) to create master index and then data list to drive script
   - 4.2. Prepare script to rename and move data files into archive structure
   - 4.3. Run script; check processing logs and manually move any problem files
   - 4.4. Archive well log data set

Appendix 2 Data Holdings Maintained by IMFO

This list comprises the main databases and datasets maintained by the IMFO Team. The list includes the significant corporate datasets, other smaller, more specialist data holdings, also maintained by IMFO, do not feature here.

Dataset	Description
Borehole Geology	Database of geological interpretations of boreholes drilled in Great Britain. These interpretations may have been supplied by the geotechnical company responsible for drilling the hole, or may have been made subsequently either by BGS geologists or other agents. Borehole Geology is heavily used in conjunction with the Single Onshore Borehole Index (SOBI) to generate 3-D models and other representations of the geology of Great Britain. Currently just over 3 million records.
BritPits	Also known as the Mines and Quarries Database, this stores information about active and inactive onshore mineral workings, also rail depots and wharves used in movement of commodities. Compiled and used by BGS to periodically produce the 'Directory of Mines and Quarries' for sale through the MineralsUK website, and for market intelligence / analysis and resource planning by internal and external customers including local councils. Data made externally available under licence. Currently approximately 140,000 records.
Coastal / Marine Database (CMD)	Database of BGS-held Marine Survey data, including an index at survey level linked to digital and analogue geophysical, sample and multibeam data and information.
Data Management Planning System (DMPS)	The system where the information on project data management is entered and stored. It is a tool for data management planning but also for capturing long-term data assets and for management of the Storage Area Network resource.
Discovery Metadata	Database holding ISO 19115 compliant metadata describing significant data holdings in BGS.
Geochemistry Database	Database for geochemical data from the land area of Great Britain and Northern Ireland (excluding groundwater). The Geochemical Baseline Survey of the Environment (G-BASE) is the most active contributor, about 60%. About 35% samples are from the Minerals Reconnaissance Programme (MRP). The remaining data is from various other BGS projects. Samples types include stream sediments, soils, stream waters, rocks, drill cores, panned concentrates and tills. There are now over 10 million analyte determinations stored in the database.

Dataset	Description
Geoscience Imagebase 2 / Geoscenic	This is the BGS database of digital geoscientific images and associated metadata, and includes images taken by BGS geoscientists as well as collections accessioned from other organisations. A subset of the images are selected and published on the BGS website as 'Geoscenic'.
Issues Log	Database and tracking system for issues identified as possible errors in corporately held data.
Lexicon of Named Rock Units	Database of definitions of rock units and other geological deposits recognised by BGS on the UK and its continental shelf. As well as being a repository of knowledge, the Lexicon is used to constrain the lithostratigraphical attribution of many other BGS databases. It is searchable on the BGS website. Currently approximately 13,200 records.
National and Global Geomagnetic database	Geomagnetic data from BGS observatories in the UK and overseas.
National and Global Seismic database	Continuous seismic waveform data for the British national seismic network for up to 12 years.
	Parametric data for national and global seismic events going back over 1000 years.
National Building Stones Database	Database of (mostly historic) buildings and construction materials. The information are collected and supplied by English Heritage staff. BGS uses the information to answer enquiries and provide advice to customers and partners, often local authorities and heritage agencies. Approximately 4000 index records plus additional related data.
National Geotechnical Properties Database	Database of geotechnical properties and other information derived from site investigation reports supplied to BGS. BGS compiles this information for use in engineering geology research, answering enquiries, and increasingly to attribute 3-D geological models. Currently approximately 81,500 index records plus additional related data.
National Groundwater Level Archive	Database of groundwater level observations from representative boreholes across the British Isles. The majority of data are collected by environmental regulators or reported by water well owners, but increasingly data from research networks is being included. Working closely with the National River Flow Archive, which is operated by the Centre for Ecology and Hydrology's national river Flow Archive, the data are used to produce monthly and annual Hydrological Summaries, and are a key data source for assessment of water resources and for analysing long term hydrological trends. Currently holds up to date time series for around 200 boreholes, historical time series for a further 1500 sites with around 50,000 sites having intermittent measurements.

Dataset	Description
National Well Record Archive/WellMaster	Database of groundwater data, compiled from the paper well record holdings of the National Well Record Archive, which receives copies of all water wells drilled in the UK, under provisions within the Water Resources Act. The digital database, WellMaster, holds geological and hydrogeological data used for hydrogeological research, answering enquiries and product development. Ancillary databases hold detailed aquifer property and groundwater chemistry data. Currently approximately 110,000 records are held, although the level of detail available is constrained by original record quality.
World and UK Mineral Statistics	Database of world mineral production statistics. Compiled and used by BGS for market intelligence / analysis, and for periodical production of 'World Mineral Production', 'European Mineral Statistics', and 'United Kingdom Minerals Yearbook' for sale or download through the MineralsUK website. Currently approximately index 241,000 records plus additional related data.

Glossary

BCL - British Coal Ltd (the transitional state-owned monopoly company prior to full privatisation)

BIRPS - British Institutions Reflection Profiling Syndicate

CA – Coal Authority (the residuary body after full privatisation, which owns the assets not transferred to the private coal operators plus all the liabilities).

CDI - Common Data Index, a Geo-Seas/SeaDataNet metadata record standard which is an enhanced ISO19115providing a metadata index to individual marine datasets.

CEH – the NERC research Centre for Ecology and Hydrology

CSR - Cruise Summary Report (previously ROSCOP), an international metadata standard for gathering cruise information.

CMD – Coastal and Marine Database,

DAC – Data Archive Centre

DBS – Database Solutions - a team under Information Management.

DCLG - Department of Communities and Local Government

DDS - NERC Data Discovery Service, a NERC web application enabling users to search a catalogue of metadata (information describing data) to discover and gain access to NERC's data holdings and information products.

DM – Data Management

DMPS Data Management Planning System – an application accessed via the BGS intranet for Project Leaders to use in planning the data management activities on their projects.

DTM - Digital Terrain Model

EDMO – European Directory of Marine Organisations.

EDMERP – European Directory of Marine Environmental Research

EIDC – Environmental Information Data Centre, the NERC data centre for the Terrestrial and Freshwater Sciences.

EOF UK Environmental Observation Framework http://www.ukeof.org.uk/

EMODNET - European Marine Observations and Data Network

G-BASE – Geochemical Baseline Survey of the Environment

GeoIndex – Web based version of the Geoscience Data Index

Geoscience Imagebase – the application that stores and enables access to the BGS collection of photographic and other images.

GIN – Geomagnetic Information Node

GSF\LaboratoryOperations – The folder on the working area of the SAN formerly used for laboratory facilities

IM Information Management – the part of the Information and Knowledge Exchange directorate dealing with the fundamental data, data structures and applications of the BGS and NGDC.

IMFO Information Management Focus and Operations- a team under Information Management.

IPR Intellectual Property Rights – Intangible property rights generated as a result of creativity for which the owner can make financial benefit and control of distribution

INTERMAGNET - *Int*ernational *R*eal-time *Mag*netic Observatory *Net*work

ISSC – Information Systems Steering Committee, BGS committee advising senior management on Information Technology and Information Systems

IUGS-CGI – International Union of Geological Sciences, Commission for Management and Application of Geoscience Information

LEXICON – The BGS Lexicon of Named Rock Units, a database designed to hold definitive and readily accessible information about all allostratigraphical, lithostratigraphical and lithodemic units of member (or equivalent) and higher rank that are used, or have been used, on BGS maps and in BGS publications

LIMS – Laboratory Information Management System, a system designed to manage data output from laboratory equipment and activities

MEDIN – Marine Environmental Data and Information Network,

MoU – Memorandum of Understanding,

MRP – Mineral Reconnaissance Programme,

MTG – The Multilingual Thesaurus of Geoscience, a thesaurus of geoscientific terms under construction by the IUGS-CGI MTG Working Group

NCB – National Coal Board

NERC Data Grid, a NERC eScience project to develop a single point of access (or portal) to NERC's data see DDS.

NGDC – National Geosciences Data Centre. The NERC data centre responsible for the earth sciences.

NGLA – National Groundwater Level Archive,

PMS – The BGS Project Management System (an application for project initiation and planning accessed via the intranet) now superseded by the RMS.

QC – Quality Control

RMS – Resource Management System. A NERC wide resource for project initiation and planning available via the Intranet.

SAN – Storage Area Network – a high-speed network of various levels of storage hierarchy that permit large volumes of data to be stored in a secure manner.

Sandbox – A piece of software that provides an 'editing area', into which data from a database can be imported, modified, then exported back to the originating database. It allows records to be edited without that editing process being visible to other users of the database.

SciFac\AG The folder on the working area of the SAN used by Analytical Geochemistry laboratory facilities

SEG-Y - Society of Exploration Geophysicists format for seismic data

SEISCANEX - EC Framework 5 Programme project to extend the accessibility and use of old paper seismic records by image archiving and digital translation to SEG-Y through the provision of Work based training and development tools.

SFC – Stratigraphic Framework Committee.

SIGMA - System for Integrated Geoscience Mapping, a structured and consistent methodology for the complete geological surveying process in BGS.

SNS –Systems and Network Support

SOBI – Single Onshore Borehole Index, the BGS digital index to records of boreholes drilled in Great Britain

UKCS – United Kingdom Continental Shelf.

UKHO – United Kingdom Hydrographic Office,

W: drive – The part of the SAN used for "working" data; projects, laboratory facilities etc. as opposed to long-term storage.

WellMaster – Hydrogeological Database holding hydrogeological information on water wells for wells and boreholes within England Wales and Scotland.

WELLOG – In-house BGS software for storing borehole logs (now superseded).

References

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

Dr. Frank Gibson and Dr. Phillip Lord, Create or Receive Scientific data, Digital Curation 101, October 6th-10th, 2008, NeSC, Edinburgh