



Managing Collections for Exploitation

Jeremy Giles

National Geoscience Data Centre

www.GeoscienceDataCentre.com

Kingsley Dunham Centre
Keyworth
Nottingham NG12 5GG
Tel 0115 936 3100

© NERC All rights reserved



Collection

- I want to start by saying what I mean by a collection
- The OED defines collect as:
 - *“bring together; assemble; accumulate; systematically seek and acquire”*
- The term is broad and describes the act of assembling objects together, along with appropriate contextual data, for an implied purpose.
- A geoscience collection is therefore:
 - *A group of geoscience objects, analogue and/or digital, that are assembled together, along with appropriate contextual data, for a specific purpose.*

Xenophanes of Colophon

(570 – 480 BC)



- “...And Xenophanes believes that once the earth was mingled with the sea, but in the course of time it became freed from moisture; and his proofs are such as these:
 - that shells are found in the midst of the land and among the mountains,
 - that in the quarries of Syracuse the imprints of a fish and of seals had been found,
 - and in Paros the imprint of an anchovy at some depth in the stone, and
 - in Melite shallow impressions of all sorts of sea products...”

<http://history.hanover.edu/texts/presoc/Xenophan.html>

© NERC All rights reserved

Geoscience Collections



- Xenophanes and his numerous successors made their observations and collections for scholarly purposes.
- The collections created by numerous scholars during The Enlightenment are the foundations on which our knowledge is built.
- However, the cost of maintaining and managing such collections is high and is constantly questioned.

© NERC All rights reserved

Geoscience Collections

www.nerc.gov.uk



- Many collection managers have been asked the naïve question:
 - *“Couldn’t we save a lot of money if we gave our collections to someone else?”*
- The simple answer of:
 - *“No, we would have to transfer resources along with the collection.”*comes as a surprise and shows the depth of analysis behind such questions.

© NERC All rights reserved

Geoscience Collections

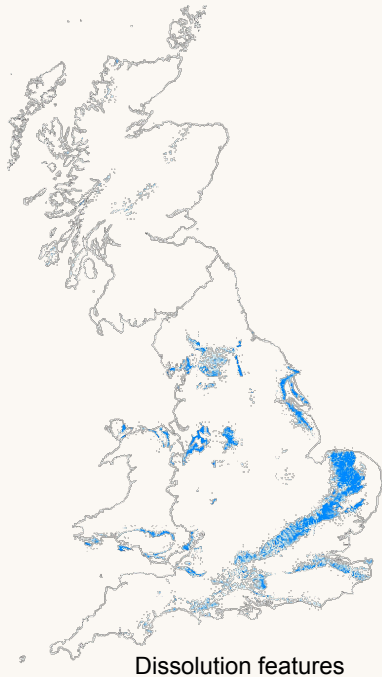
www.nerc.gov.uk



- Geoscience collections need to justify their societal value by contributing to the development of products and services that do at least one of the following:
 - Create wealth;
 - Reduce risk;
 - Improve quality of life; or
 - Improve quality of the environment

© NERC All rights reserved

Geoscience Collections



- Over the past few years BGS has been developing a range of products and services based on its analogue and digital collections.
- This has only been possible because the BGS analogue and digital collections have been managed as an asset with the intension of exploiting them.
- This is in contrast to some organisations that treat their collections as a liability.

© NERC All rights reserved

Managing Collections for Exploitation

- Define a collections management strategy;
- Develop a collections management framework covering all aspects of the collections;
- Clearly define and document what comprises an individual collection;
- Establish a clearly defined vocabulary to be used in the abstraction of information during database entry;

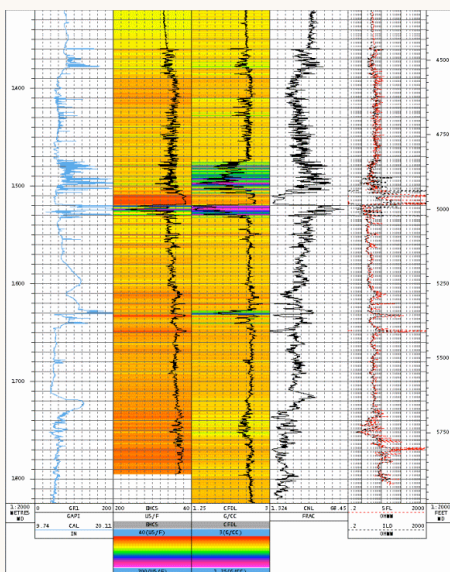
© NERC All rights reserved

Managing Collections for Exploitation

- Create, maintain and disseminate appropriate metadata for the collections;
- Create, maintain and disseminate digital indexes to the collections;
- Understand and document the limitations of the information contained in collections; and
- Use quality management techniques to produce standards, best practice, collection management plans, and specific procedures and work instructions.

© NERC All rights reserved

In Simple Terms



- Plan what you intend to do and communicate it;
- Design the structure and systems to deliver the plans;
- Understand the individual collections, why they were collected and how they can be used, re-used and re-purposed;
- Use metadata as a collections asset register;

© NERC All rights reserved

In Simple Terms

www.bgs.ac.uk

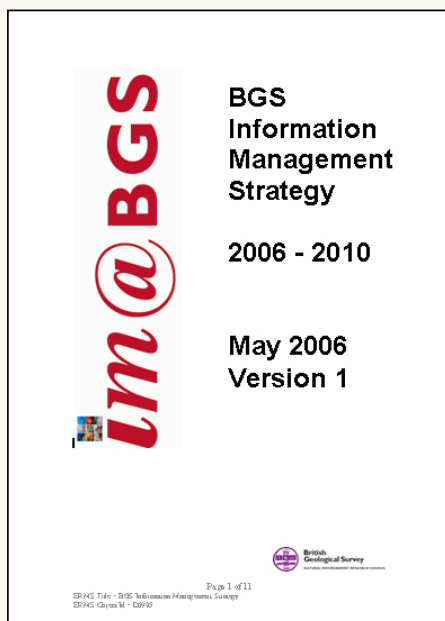


- Use spatial indexes to manage and publicize collections;
- Manage the lists of terms used to attribute a collection;
- Know the limitations of the collections; and
- Work to improve quality.

© NERC All rights reserved

Plan and Communicate

www.bgs.ac.uk



- The name of the document may change from place to place but the purpose is clear.
- There is a requirement for a single overarching document that lays out the way in which collections are going to be managed and developed over the medium term.

© NERC All rights reserved

BGS Strategy Headlines

- Manage BGS Collections in accordance with all relevant legal and policy obligations;
- Manage BGS Collections as an asset;
- Build user-confidence by implementing and maintaining a quality management system;
- Create, maintain and disseminate metadata;
- Promote rapid discovery of individual items within collections of all types using spatial indexes; and
- Provide tools that enable geoscientists to use collections with confidence.

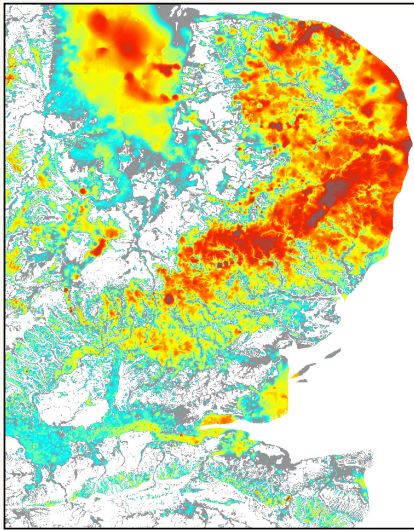
© NERC All rights reserved

Structure and Systems

- A strategy needs to be implemented.
- This will be done by a range of people with a range of skills and expertise.
- Roles and responsibilities need to be assigned.
- Resources need to be sought, allocated and accounted for.
- Projects need to be started, plans made and work done.
- Policies, procedures and work instructions need to be written.

© NERC All rights reserved

Thickness of Superficial Deposits



Thickness of superficial deposits

- As an example through the following slides I am going to use the BGS Borehole Records Collection.
- This was one of the key collections used to produce the BGS Thickness of Superficial Deposits map.

© NERC All rights reserved

Know the Individual Collections

- A geoscience collection is a group of objects, analogue and/or digital, that are assembled together, along with appropriate contextual data, for a specific purpose.
- I think that the purpose is key to recognising an individual collection.
- The purpose may be that the individual items are:
 - Beautiful mineral specimens
 - Type fossils used to define a species
 - Digital geophysical logs used in hydrocarbon exploration.

© NERC All rights reserved

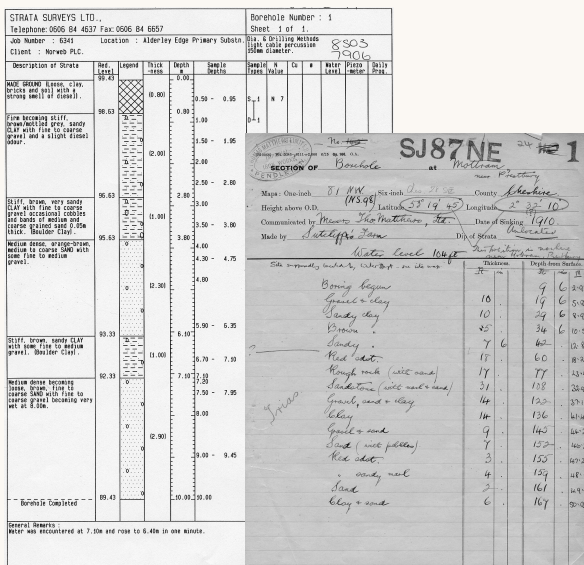
BGS Borehole Records Collection



- Geologists are interested in the third dimension when creating geological maps.
- Boreholes drilled by industry are a valuable source of such data.
- In Britain various laws allow the geological survey to have copies of mineral exploration and water boreholes.

© NERC All rights reserved

BGS Borehole Records Collection



- Over the course over of 170 years a collection of 1.2 million paper borehole logs has been assembled to support geological mapping.
- These paper logs were scanned in 2001 and 2002 and digital facsimiles created in TIFF format.

© NERC All rights reserved

Metadata

www.nerc.gov.uk

- Metadata is a tool for describing collections and then publishing that description through one or more metadata portals.
- It is most useful as an asset register of the collections.
- It aids discovery and promotes re-use and re-purposing.



© NERC All rights reserved

Discovery Metadata Entry BGS Borehole Records Collection

www.nerc.gov.uk

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

BGS Discovery Metadata Dataset
Borehole Records Collection.

Dataset description
Records of all onshore (or near shore) boreholes, trial pits, shafts and wells held in the BGS archives in either paper, microfilm or digital format. The records range from simple single page lithological logs through to hydrocarbon completion reports. Spatial coverage will vary considerably depending on drilling activity, collecting activity and donations. The majority of new data is from site investigation reports with concentrations in urban areas and along transport routes. Current collection over 1million records covering the whole of Great Britain with 50,000 new records added per annum. Some records date back to 1600 but the majority date from 1900 onwards.

Constraints
Some borehole records have been deposited with commercial-in-confidence restrictions. In this case the position of most boreholes can be given but no other Index Information such as depth. BGS standard conditions of use apply and any copyright constraints.

For more information please contact :
Enquiries
British Geological Survey
Keyworth
Nottingham
NG12 5GG
Tel : +44 (0)115 936 3143
Fax : +44 (0)115 936 3276
Email :enquiries@bgs.ac.uk

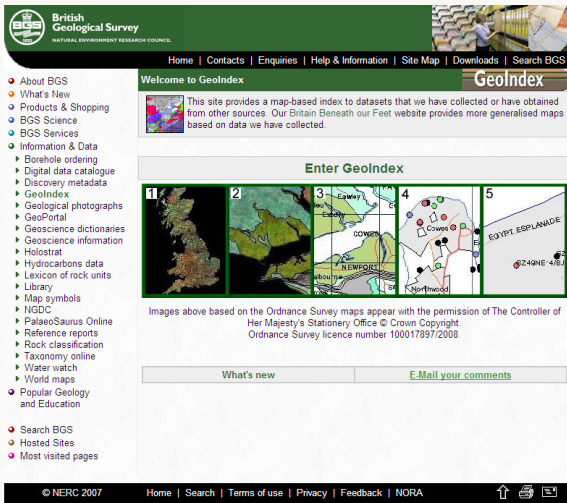
Associated dataset(s)
Scanned images of the BGS Borehole Records Collection.
Single Onshore Borehole Index

Location
W -8.65 60.84 N
S 49.77 2 E

Reviewed 15th November 2006

© NERC All rights reserved

Use Spatial Indexes



- Geoscience collections are normally geospatial.
- Individual items are collected from a specific location and that location is critical to the context of the item.
- Users of collections normally want to find items by spatial searching; e.g. trilobites of Shropshire or gamma logs from boreholes in the Piper Field in the North Sea.

© NERC All rights reserved

BGS Borehole Records Collection Spatial Index

Zoom To	REFERENCE	NAME	GRID_REF	EASTING	NORTHING	PRECISION	LENGTH	YEAR_KNOWN	SITEREPORT	HELD_AT	ID
1	TQ28SE162	BURLINGTON HOUSE WESTMINSTER	TQ 29219 80530	529219	0180530	± METRE	137.16			KW	591646

© NERC All rights reserved

Manage Vocabulary

www.bgs.ac.uk

- Exploiting analogue collections, such as the BGS Borehole Records Collection, requires databases to be populated.
- The permitted values or codes used in given fields need to be carefully managed.
- Terms that are to be used need to be clearly defined and unambiguous.

© NERC All rights reserved

Manage Vocabulary

www.bgs.ac.uk

Code	Description	Translation	Status
char(2)	varchar(60)	varchar(50)	char(1)
Assc Desc	Assc Desc	Assc Desc	Assc Desc
*	Base Of Bed		U
AC	Angular Unconformity		U
DC	Disconformity		U
DR	Rockhead Uncertain, Obsolete Code Use Ru Or Ra Instead		U
ES	Erosion Surface		U
FT	Fault		U
GR	Gradational		U
HB	Obsolete Code For Total Depth Of Bhole Use Td Instead		U
PA	Passage By Alternation		U
PC	Parallel Unconformity		U
RA	Approximate Position Of Rockhead		U
RH	Position Of Rockhead		U
RJ	Used Where Lith Code Link Includes Both Solid Rock & Drift		U
TD	Total Depth Of Borehole		U
UC	Unconformity (No Longer Used For Rockhead - Use Rh, Ra Or Ru)		U

- Care must be taken when adding new values to “List of Values”.
- The definition of additional terms must not conflict or overlap with existing terms.
- Otherwise the meaning of the field values will become confused and hence devalued.

© NERC All rights reserved

Know the Limitations

www.nerc.ac.uk



- No collection is perfect.
- A choice has been made to add some items to the collection and to exclude others.
- Such selection causes a sampling bias.
- If the collection is to be exploited, especially for a use not envisaged at collection, such limitations must be understood.

© NERC All rights reserved

Know the Limitations

www.nerc.ac.uk

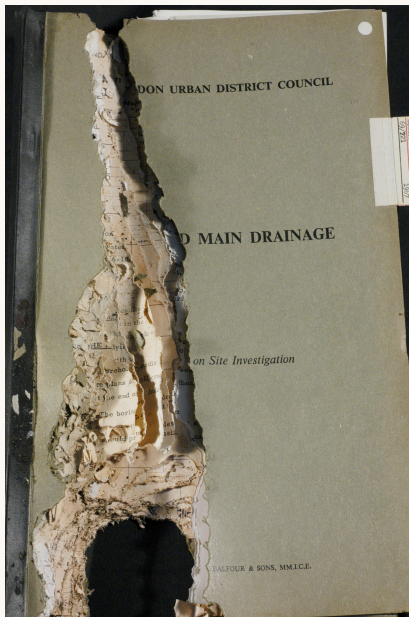


- Intellectual property is a common limitation.
- This is principally because long standing collections acquired many of their items prior to the development of the concept.
- There are ways around this problem but it requires planning, time and resources to complete.

© NERC All rights reserved

Know the Limitations

www.bgs.ac.uk

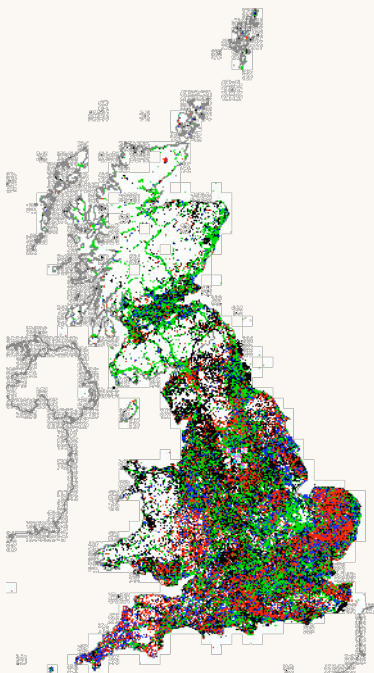


- Other limitations include:
 - Incomplete contextual data
 - Missing items
 - Misplaced items
 - Poor preservation

© NERC All rights reserved

Limitations of the BGS Borehole Records Collection

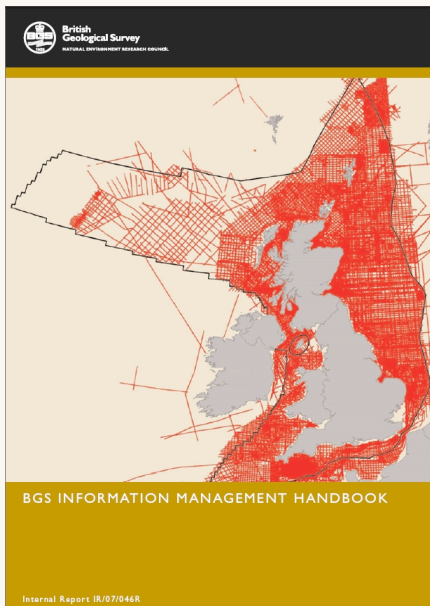
www.bgs.ac.uk



- Unrepresentative distribution
 - urban areas, roads, ground water resource and mineral resource areas are well represented
 - Rural areas are under represented.
- The majority of records are by voluntary deposit, some companies never volunteer records.

© NERC All rights reserved

Work to Improve Quality



- Quality can be improved progressively
 - Produce clear policy
 - Identify, document and maintain best practice
 - Establish clear procedures for activities
 - Use work instructions
 - Audit

© NERC All rights reserved

Work to Improve Quality

Document Business Rules and implement in databases

No	Complete ²	Priority ³	Rationale / Discussion / Technical ⁴⁵⁶	Result
2	Yes	High	JRAG: Hydrocarbon Boreholes as indicated by the PURPOSE attribute should have a DRILLED_LENGTH greater than 150 metres	<p>12 Oct 2006 Action KAH to check current data and provide a list to RCB and RJGI to make corrections.</p> <p>Completed 13 Dec 2006 (KAH) – see SOBIRule2 ... maybe the business rule wrong – perhaps assuming hydrocarbons = oil</p> <p>12 Oct 2006 Action RCB and RJGI to make corrections or decide to continue action or not</p> <p>12 Oct 2006 Action KAH To implement a check constraint to prevent subsequent occurrences. Data entry staff to be informed by notifying team leaders of this actions spreadsheet.</p> <p>10 Jul 2007 All actions dropped. Meeting decided we cannot identify Hydrocarbons bores in the database with sufficient certainty to implement this</p>
3		Medium	JRAG proposal Combined with 12	Combined with 12
4	Yes	High	JRAG: Site Investigation Boreholes as indicated by the PURPOSE attribute should have a DRILLED_LENGTH less than 150 metres KAH: How are Site Investigations defined in SOBI? KAH: Best to use “<=” rather than “<” unless there’s a good reason for using “<” ... it’s just that “<” is often taken to mean “<=”	12 Oct 2006 Meeting decided not appropriate. No action taken.

© NERC All rights reserved

Work to Improve Quality

Use simple Issue Logs to capture errors from users of the collections


#2_862 Scanned document error	S1873E147 - first page not scanned only page 2	New	High		Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADNg	08/05/2008 ADJavel
#2_861 Scanned document error	B0510-1009628 - Other - This page 1030/1030 but only has 1020 attached	New	High		Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADUmbo	08/05/2008 ADJavel
#2_868 Scanned document error	B0510-202851 - Other - page does not relate to this record, wrongly indexed	New	High	delete scan	Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADrapa	08/05/2008 ADJavel
#2_867 Scanned document error	B0510-202851 - Other - page relates to a different record, wrongly indexed	New	High	Delete this page	Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADrapa	08/05/2008 ADJavel
#2_863 Scanned document error	B0510-185036 - Page Missing - scan related to another record	New	High	Delete scan	Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADrapa	08/05/2008 ADJavel
#2_862 Scanned document error	B0510-185036 - Other - scan relates to another record	New	High	Delete scan	Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADrapa	08/05/2008 ADJavel
#2_861 Scanned document error	B0510-185036 - Other - wrong scan indexed, belongs to another record	New	High	Delete scan	Red Boxes To Action	08/05/2008 Alan Weller	22/04/2008 ADrapa	08/05/2008 ADJavel
#2_849 Scanned document error	B0510-185036 - Other - Wrong scan indexed	New	High	Delete page	Assigned to Red Boxes for action	08/05/2008 Alan Weller	22/04/2008 ADrapa	08/05/2008 ADJavel
#2_848 Scanned document error	B0510-303750 - Page Quality - ILLIBLE WRITING	New	High		Assigned to Red Boxes for action	08/05/2008 Alan Weller	17/04/2008 ADAchid	08/05/2008 ADJavel

Issue Log with entries for scans of BGS Borehole Records Collection

© NERC All rights reserved

Work to Improve Quality

Use automatic error trapping on digital collections

BGS Technical Metadata Application				
	Alerts Found: 37			
	Alert	Manager	Category	Clump
Databanks Objects (Tables, Views, Privileges, etc...) Data & Application Managers Projects Oracle Users Glossary Alerts Schemas Documentation Business Rules	101	Martin L Nayembil	Data Error	Data Error
	102	Martin L Nayembil	Data Error	Data Error
	103	Martin L Nayembil	Data Error	Data Error
	104	Martin L Nayembil	Data Error	Data Error - Ascii
	105	Martin L Nayembil	Data Error	Data Error - Ascii
	106	Martin L Nayembil	Data Error	Data Error - Ascii
	107	Martin L Nayembil		
	108	Martin L Nayembil		
	109	Martin L Nayembil		
	110	Martin L Nayembil		
	110	Rod Owen		
	111	Martin L Nayembil		
	111	Martin L Nayembil		
111	Rod Owen			
111	Rod Owen			
112	Martin L Nayembil			
112	Rod Owen			
113	Martin L Nayembil			
113	Rod Owen			

Alert Number:	105
Alert Description:	Ascii(10) found in Keyworth text field(s)
Alert Manager(s):	Martin L Nayembil
Alert Category(s):	Data Error
Alert Clump(s):	Data Error - Ascii
Health Level (%):	-337.931034483
Alert Date:	2008-03-27 04:24:33.0
Alert Count:	127
Alert Reference:	29
Log File:	Open Alert Log File

© NERC All rights reserved

Conclusion

- Managing collections for exploitation requires a different approach.
- Collections need to be recognised as an organisational asset which can potentially be exploited in the development of products and services.
- Once collections are linked to specific income generating products and services their strong scientific justification is also supported by a robust business case.
- This is achieved though improving the way that collections are managed.

Conclusion

- Plan what you intend to do and communicate it;
- Design the structure and systems to deliver the plans;
- Understand the individual collections, why they were collected and how they can be used, re-used and re-purposed;
- Use metadata as a collections asset register;
- Use spatial indexes to manage and publicize collections;
- Manage the lists of terms used to attribute a collection;
- Know the limitations of the collections; and
- Work to improve quality.

Questions

www.bgs.ac.uk

- Web addresses
 - www.bgs.ac.uk
 - www.bgs.ac.uk/geoindex
 - www.bgs.ac.uk/discoverymetadata
 - www.GeoscienceDataCentre.com
 - www.geologyshop.com
- E-mail address
 - enquires@bgs.ac.uk

