



British
Geological Survey

NATURAL ENVIRONMENT RESEARCH COUNCIL

Gateway to the Earth

National Geoscience Data Centre (NGDC): Drivers, experiences and lessons from seeking CoreTrustSeal certification

Garry R. Baker

Head of National Geoscience Data Centre (NGDC)

grba@bgs.ac.uk

NGDC

- One of NERC Environmental Data Centres

- National Geoscience Data Centre (NGDC)
- Centre for Environmental Data Analysis (CEDA)
- Environmental Information Data Centre (EIDC)
- Polar Data Centre (PDC)
- British Oceanographic Data Centre (BODC)

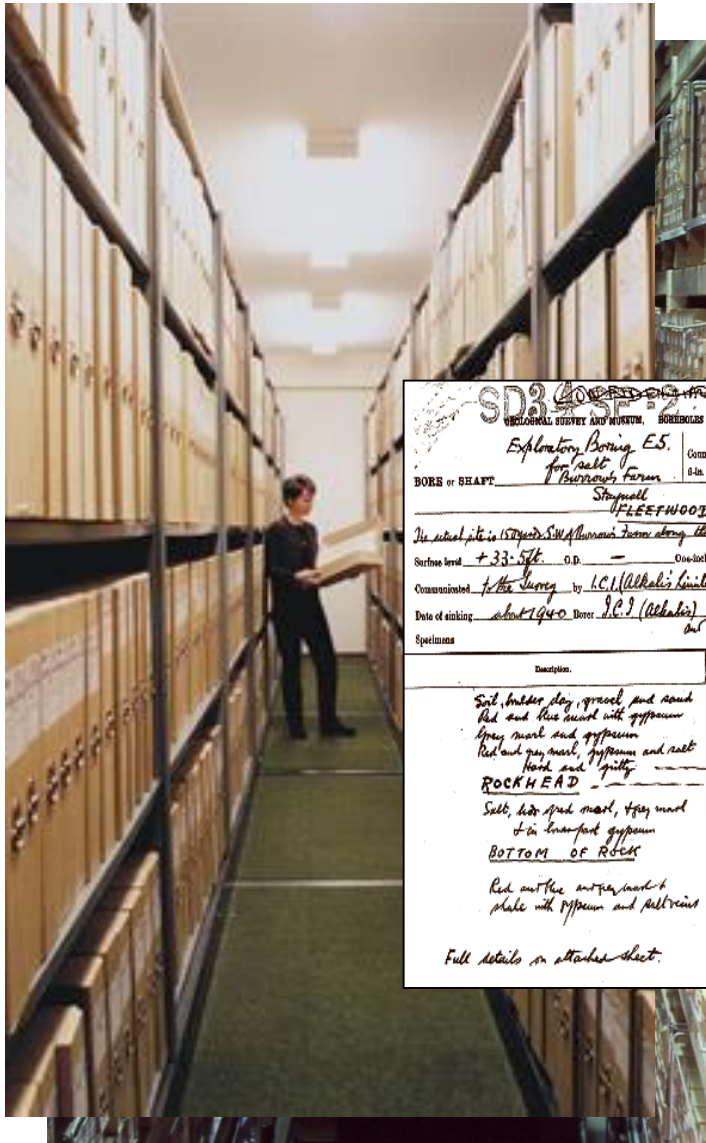
- Manage environmental scientific data and information from a range of differing geological or geoscience disciplines

- Holds historical/legacy observation data up to real-time geological processes, sensor networks and data streams

- Preserve this data for the long-term as evidence of existing of the current scientific/research record and to make it easily accessible for future re-use



NGDC: Breadth of Data and Information



SD 240 10
VICTORIAN SURVEY AND MUSEUM. BOREHOLES AND SHAFTS. 43NW
Exploratory Boring, E.S. for salt of Burrows' Farm County Gloucestershire
 BORE or SHAFT *Stapwell* 8-in. Map 4-3NW **9**
FLEEFWOOD GSM

The actual pit is 150 feet SW of Burrows' Farm along the banks of the lagoon.
 Surface level *+33.5 ft.* O.D. *—* One-inch Map 66 (*Blackwood*)
 Commenced *1/2 the Survey* by *L.C.I. (Alkali's Limited), Northwick & Richard*
 Date of sinking *about 1900* Donee *J.C.I. (Alkali's) Hestonwood*
 in name of *Shon & Co.*

Description.	INDICONS.		FEET.	
	Rev.	Tacks.	Dist.	Totals
<i>Soft, bluish clay, gravel, and sand</i>	<i>158</i>	<i>0</i>	<i>158</i>	<i>0</i>
<i>Red and blue sand with gypsum</i>	<i>6</i>	<i>0</i>	<i>273</i>	<i>6</i>
<i>Grey marl and gypsum</i>	<i>45</i>	<i>6</i>	<i>325</i>	<i>0</i>
<i>Red and grey marl, gypsum and sand</i>				
<i>Hard and gritty</i>				
ROCKHEAD				
<i>Salt, bit. sand, marl, grey marl</i>				
<i>& in lower part gypsum</i>				
BOTTOM OF ROCK				
<i>Red on the surface, marl & shale with gypsum and salt veins</i>				

Full details on attached sheet.

Depth	Type	Properties			SPT	D ₅₀
		Strength	w	M		
0.30	0					
0.30-1.00	10					
1.30	0					
1.50-2.00	10/15					
2.30	0					
2.50-3.00	10/15					
3.30	0					
3.50-4.00	10					
4.30	0					
4.50-5.00	10					
5.30	0					
5.50-6.00	10					



Drilling		Ground Water								
Type	From To	Size	Placed	Shock	Behaviour	Gravel	Date	How	Case	Filter
Shell and Pipe	0.1 - 8.0	8.0	W.C.	0.3	Slight seepage		29.2.85	MS	Nil	Nil
							29.7.95	0.0	Nil	Dry

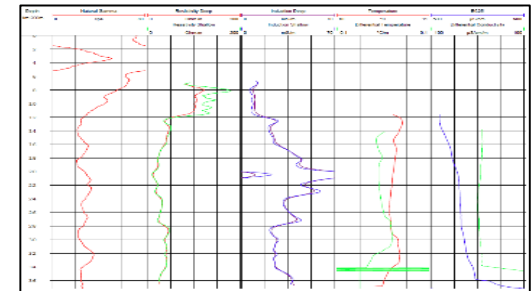
Remarks

Borehole Record	Project Silver Magnet and Partners Denny Potteries, Denny	Contract S487
exploration associates		Borehole 1 Sheet 1 of 1

NGDC: Data Holdings

- Complex, diverse range of environmental / geoscience datasets including:
 - **900TB+ on SAN** (plus another 500TB in tape archive)
 - Oracle RDBMS with over 3000 objects
 - Near-real time data from sensors on monitoring sites and observatories
 - Over **1.5 million open access borehole records** with **3.7 million associated scanned images**
 - **500,000+ scanned images** containing site specific geological information
 - **200,000+ digital geophysical well data logs and curves**
 - **150,000+ photographs and imagery** (core photos, 3D fossil scans)
 - **50,000+ spatial data files (GIS)**
 - **Business intelligence - logs of app usage, queries upon data stores, social media.....**
 - **Data Warehouse built from sensor monitoring sites, sub-surface parameter data**
 -plus....**450km of core, millions of samples and 17 linear KM of paper records** (Many differing analogue data types)

Level (mOD)	Depth (m) (Thickness)	Description	Legend	Water
	(0.80)	MADE GROUND : Angular medium to coarse GRAVEL of Geotextile layer at 0.10m and 0.20m bgl. granite.		
46.86	0.80 (0.50)	MADE GROUND : Stiff clayey very sandy fine to medium, occasionally coarse GRAVEL. Gravel is subangular limestone and flint.		
46.38	1.30 (0.35)	Firm friable light brown slightly gravelly CLAY. Gravel is subrounded fine flint. (HYTHE FORMATION)		
46.03	1.65 (1.35)	Firm occasionally laminated green-grey mottled brown sandy CLAY. (HYTHE FORMATION)		



Drivers

- It's becoming a **crowded marketplace**, there's a need to stand out from the crowd, **demonstrate our professionalism at longer-term data & information management** (help demonstrate the value and benefits of the repository)
- We need to **consider the full lifecycle of the repository** from the core underpinning infrastructure, staff, policies, procedures, workflows and systems; starting from best practice guidance and donation through to delivery and ultimately re-use
- **Benchmarking our processes and service** against a recognised standard framework in CoreTrustSeal, provides evidence of the level of quality, compliance and impact from the repository
- The need to **provide confidence to our funders** on the investment they make within our repository
- **Alignment to scientific journals and publications** – they now expect to link to the data underpinning the scientific paper but only in reputable/certified repositories

Experiences

- Getting the team together was invaluable, proved a superb vehicle to enhance communications, discuss new ideas and agree priorities
- We had a range of policies, procedures and systems in place for engagement, donation, management and delivery
 - Some were robust and well thought-out
 - Some were okay but lacked clarity/not easily understood, or a little outdated
 - Some were implied, “knowledge in heads”
- Our community/global standards compliance is already strong, the CoreTrustSeal framework reinforced this
- We do have a diverse range of user types and communities (for both donation and data re-use) which presents challenges
- Clear reminder of the breadth of a modern data repository

Presence of website	as needed
Start of website	2013
End of website	Not known
Organisation	Research
Organisational Unit	British Geological Survey
Address	Environmental Science Centre, Madingley Hill, Norwich
City	Norfolk
Country	United Kingdom
Postal Code	NR6 3BG
E-mail	enquiries@bgs.ac.uk
Telephone	+44 (0)1509 506 3143
Fax	+44 (0)1509 506 3076
Keywords	Research
High Impact	geoscientificInformation (Information pertaining to earth sciences)
Keywords and Equivalents	DATA 2013/04/04 DOI 10.1016/j.data 10.1016/j.data 10.1016/j.data
Initial Release Date	2008-2009 / British National Grid (2008)
Geographic Location	North boundary: 63.00 West boundary: -1.00 East boundary: -0.75 West boundary: -0.80
Metadata Language	English
Metadata last updated	1st November 2013

NGDC Data Remit/Scope

The National Geoscience Data Centre (NGDC) is hosted by the British Geological Survey (BGS) and delivered on behalf of the Natural Environment Research Council (NERC). It is responsible for managing geoscientifically valuable information and datasets. Data held by NGDC covers many geological disciplines that deal with the physical structure of the Earth and the processes which act on it as measured from geological time to near-real time sensor and data streams. It is the NGDC policy to preserve the existing scientific record and make it accessible for future re-use to a wide range of user communities.

NGDC Ingestion Policy

It is the policy of NGDC to store data that is:

- high quality, scientifically significant and of potential research or commercial interest to scientists and academics or policy makers in the future
- a geoscientific data output from NERC funded earth science projects and programmes
- digitally born, enabling easy future re-use and access by a diverse range of end-user communities
- deposited by the owner or someone who has the rights to donate the data to NGDC (therefore granting NGDC a non-exclusive “in-perpetuity licence” to hold and disseminate the data)
- compliant with our data policy and data value checklist.

It is also the statutory responsibility of NGDC to receive data associated with:

- boreholes drilled from the surface to a depth of more than 30m for mineral exploration/extraction purposes (Mining Industry Act 1926)
- boreholes drilled deeper than 13 m for water (Water Resources Act 1991 and Water (Scotland) Act 1980).
- Hydrocarbon well data (as per POM/95). See [National Hydrocarbons Data Archive \(NHDA\)](#).

It is the policy of NGDC to ensure data is deposited with the following:

- standards compliant metadata that explains the dataset and makes it easy to discover and facilitates re-use
- terms and conditions that clearly define how the data can be used in the future (NGDC prefer to hold data with no access or user restrictions - “open data”)

It is the policy of NGDC to:

- store a copy of the data “as supplied” in a Data Holding Store (in an agreed format)
- incorporate (upon landing permits) the data within our existing nationally consistent datasets
- provide data management advice and guidance to geoscientists working upon earth science projects
- ensure the data is discoverable and provide open access to any unrestricted data held in the Data Holding Store or the nationally consistent datasets
- create Digital Object Identifiers (DOI’s) for appropriate digital datasets on request, facilitating data citation and display these alongside the data
- encourage the use of these datasets for a full range of academic and commercial purposes including future scientific

Checklist

Essential criteria: These are legal or regulatory criteria and answering “Yes” to one or more of the questions below will automatically result in selection for retention.

Legal/statutory considerations	Yes	No
Is there a legal or legislative reason for NERC to retain the data under any of the following:		
• Science & Technology Act 1965		
• Mining Industry Act (1926)		
• Water Resources Act (1991)		
• Petroleum Operations Notice (PON 9) regulations (on-shore and off-shore)		
• Public Records Act (1958 & 1967)		
Has or could the data been used in litigation, public enquiries, police investigations or any report or paper that could be legally challenged?		
Are there any financial or contractual obligations that require us to retain the data?		

Important criteria: These are primary criteria and answering “Yes” to at least one of the questions from each section below should result in selection for retention.

Policy	Yes	No
Does the NGDC Data Policy apply to this data?		
Are the data a result of NERC/BGS funded activities?		
Does this data fall within the NGDC remit?		
Scientific or historic value		
Does the data have a geographical or temporal extent that makes it useful to others?		
Does the data have historic value i.e. does it represent a landmark in scientific discovery?		
Do the data include changes in processing methods, new standards or set any precedents?		
Do the data support current projects or trends in science?		
Is there likely to be further work in this or associated sciences?		
Are the data likely to meet the future needs/direction of the scientific community?		
Do the data contribute to a wider collection?		
Is there potential for re-use of the data?		
Are the data cited in a publication?		

Search results

New Search | Refine Search

Sighthill Phase 2 Ground Investigation

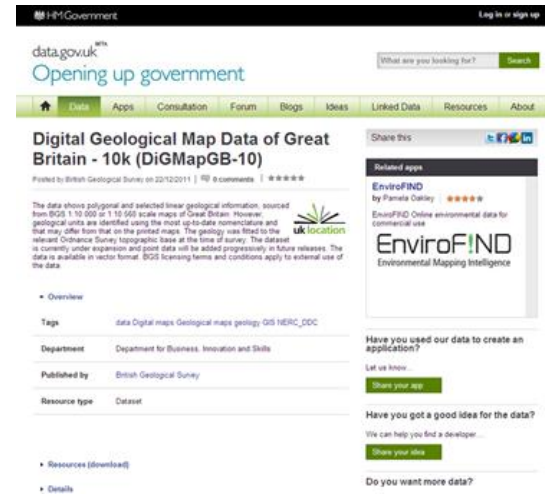
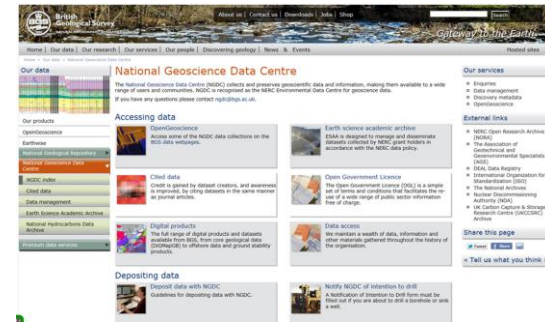
Site: Sighthill Phase 2 Ground Investigation
 Name: Sighthill Phase 2 Ground Investigation
 Address: Sighthill Phase 2 Ground Investigation
 Project: Sighthill Phase 2 Ground Investigation
 Date: Sighthill Phase 2 Ground Investigation
 Coordinates: Sighthill Phase 2 Ground Investigation
 Data File available: Sighthill Phase 2 Ground Investigation

*Please note some search results may be hidden. To see all search results, click on the 'Show all search results' link.

© Crown Copyright. All Rights Reserved. Ordnance Survey. Licence No. 100019200. Ordnance Survey is a registered trademark of Ordnance Survey.

Lessons Learnt - 1

- Reminded us of the importance of keeping updated, accessible documentation
- Process of seeking CoreTrustSeal accreditation brought the NGDC team closer together, it created time for sharing knowledge, experiences and ideas (led to better communications and understanding)
- We identified some areas that needed improvement this included processes, workflows, clarity on roles, prioritising some future work activities and enhancements to our systems
 - *Data preservation, Disaster recovery, Terms & Conditions, access and use*
- Provides a vehicle to explain the business needs of the repository to managers and funders, especially as we undertake the continuous professional development aspects of the framework
- Encouraged us to enhance our expectations (requirements) from geoscience/environmental data donators
- Encouraged us to look more seriously at future stakeholder engagement activities and plans



Acceptable digital formats

Data should normally be provided in a non-proprietary format (for example .csv rather than a Microsoft Excel spreadsheet).

The following formats are accepted by the NGDC:

Data type	Preferred format	File extension
Geotechnical data	Association of Geotechnical and Geo-environmental Specialists (preferably version 3.1 or 4.0)	.ags
Geophysical data	Log ASCII Standard	.las
	Seismic data	.sdy
Generic scientific data	Sidescan sonar data	.xtf
	Microsoft Excel files	.xls, .xlsx
	Comma-separated value files	.csv
	Data files (with readme file on software)	.dat
	Tab delimited data file	.pdf
Text	Portable document format (PDF, PDF/A)	.pdf
	Extensible mark-up language	.xml, .json
	Microsoft Word document	.doc, .docx
Presentations	Text file (plain/ASCII)	.txt
	Rich text format	.rtf
Databases	Microsoft Powerpoint presentation	.ppt
	ESRI shapefile/MapInfo/QGIS files/GeoTIFF	
Databases	Microsoft Access database	.accdb
	Oracle export MySQL export format	

Lessons Learnt - 2

- Both donators, funders and journal publishers really do like the robustness and reputation of certified repositories;
 - *NERC considering specifying CoreTrustSeal for their Data Service (as part of future commissioning processes)*
 - *BGS highlight the certification in discussions / presentations concerning their high-profile science facilities and observatories*
 - *Appreciated by key donators reinforcing the professionalism, quality, robustness, long-standing, public good aspects of the data centre*
- The role of training in disseminating best practice in data/info management and preservation to research or repository staff plus other external stakeholders (helps to mitigate future risks)
- Evidence generated from the CoreTrustSeal framework provides a sound platform from which its easier to assess and collaborate on FAIR, INSPIRE and other global data initiatives

