



## **Implementation of the CoreTrustSeal**

The CoreTrustSeal board hereby confirms that the Trusted Digital repository National Geoscience Data Centre (NGDC) complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.

The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on January 24, 2018.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

## Assessment Information

Guidelines Version:	2017-2019   November 10, 2016
Guidelines Information Booklet:	<a href="#">CTS Requirements 2017-2019</a>
All Guidelines Documentation:	<a href="#">Documentation</a>
Repository:	National Geoscience Data Centre (NGDC)
Seal Acquiry Date:	Jan. 24, 2018
For the latest version of the awarded CoreTrustSeal for this repository :	<a href="https://www.coretrustseal.org/why-certification/certified-repositories/">https://www.coretrustseal.org/why-certification/certified-repositories/</a>
Previously Acquired Seals:	None
This repository is owned by:	<b>British Geological Survey</b>
	UK
	T 0115 936 3100
	E <a href="mailto:ngdc@bgs.ac.uk">ngdc@bgs.ac.uk</a>
	W <a href="http://www.bgs.ac.uk/">http://www.bgs.ac.uk/</a>

# Assessment

## 0. Context

### Applicant Entry

*Self-assessment statement:*

Repository Type: National repository system, including governmental

The National Geoscience Data Centre (NGDC) is the UK national repository for geoscience data. It is provided by and co-located with the British Geological Survey's (BGS) Informatics Directorate. The NGDC/BGS is funded via its parent organisation the Natural Environment Research Council (NERC).

The NGDC is one of the five NERC Environmental Data Centres (EDCs) funded to provide data centre functions and services across the range of scientific communities funded within the research council.

The NGDC holds geoscience data assets primarily from NERC funded geoscience/Earth science research grants and programmes as well as those created by the Survey's own scientific programmes, or received under statute. This includes an increasingly wide variety of data from global geoscience projects, collaborations and initiatives.

Brief Description of the repositories designated community:

The designated community for the NGDC consists of a wide range of users of subsurface geoscience data and models. Users are typically in academia, local authority organisations, and industry e.g. hydrocarbons industry, environmental consultants, as well as the geotechnical and site investigation sector.

Level of Curation Performed

D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy.

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

The goal of the NGDC is the long-term professional preservation and dissemination of its data assets holdings. The data forms both an evidence base for existing and current scientific programmes, underpins existing scientific products and is as a source of critically important scientific data for inclusion in future science projects, programmes, applications, systems, products or decision support systems. The NGDC facilitates scientific peer-review processes, open access and re-use as well as the generation of impact for the wider UK industry and businesses.

The NGDC has been in existence since before 2000 but its functions have been provided to a greater or lesser degree within the BGS since its encapsulation within the Natural Research Council (NERC) in 1965.

The British Geological Survey has been in existence since 1835, under a number of organisational titles and is one of the oldest geological surveys in the world. However, the data (both analogue and digital) held by the data centre has been derived from more than 200 years of geological/geoscience projects and programmes.

The vast majority of the NERC funded and statutory data is openly available for future utilisation in accordance with the UK government 'Open Government Licence (OGL) with no barriers to re-use by external users or communities. Some value-added/interpreted datasets or information products are also held and treated as licensed products, which may be made freely available or incur small charges to license and access them depending on the status of the end-user (academic, commercial etc.).

The primary web page for the NGDC can be found at <http://www.bgs.ac.uk/services/ngdc/home.html>

The NGDC has a range of policies that underpin the functions and services it delivers, including the NERC Data Policy, metadata, collections, data management planning, digital preservation and preferred formats.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 1. Mission/Scope

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The Natural Environment Research Council (NERC) and the British Geological Survey (BGS), through its strategy (<http://www.nerc.ac.uk/about/whatwedo/strategy/>), Data Policy (<http://www.nerc.ac.uk/research/sites/data/policy/>), Ingestion Policy (<http://www.bgs.ac.uk/services/ngdc/remit.html>) and Digital Preservation Policy (<http://www.bgs.ac.uk/downloads/start.cfm?id=3173>), encapsulate the value of scientific data and the importance attached to the long-term professional management and preservation of the data assets as both an evidence base of existing scientific projects and for future re-use.

The National Geoscience Data Centre (NGDC) mission is inherited from its requirement to hold:

Statutory data sets as outlined in Acts of Parliament, model clauses and guidance documents over many years (as outlined in the entitlement to geoscientific data in the Geological Survey Act 1845, Petroleum Act 1998 (1), The Mines and Quarries Act 1954 145 (1e), Science and Technology Act 1965 and Water Resources Act 1991 (198 & 205).)

•

Scientific programme outputs from the Survey (both in the UK and globally)

•

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

NERC funded 'geoscience/Earth science' data generated from the many NERC-funded grants commissioned each year

- 

The NGDC's role as the national geoscience data centre assumes an indefinite retention of the data in its care. The present RCUK data policy states that all research data should be retained for 10 years after it was last used.

The NGDC is embedded within the Informatics Directorate of the BGS, with the aim of efficiently aligning these operational requirements within a single data centre. The remit of the NGDC including a detailed statement for the designated community regarding the repository policy on preservation and long-term reuse of data is available at <http://www.bgs.ac.uk/services/ngdc/guidelines.html>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 2. Licenses

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

To encourage the use and re-use of data the NGDC makes data freely available wherever possible via the BGS website at <http://www.bgs.ac.uk/opengeoscience/home.html?src=topNav>. Data are released using an Open Government License <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/> with the terms of use and form of any acknowledgement included.

As a Public Sector Information Holder (PSIH), BGS ensures that information is available on the terms and conditions that reflect the principles of the Information Fair Trader Scheme (IFTS) <http://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/ifts-and-regulation/>

The NGDC complies with NERC Data Licensing and Charging Policy <http://www.nerc.ac.uk/research/sites/data/policy/nerc-licensing-charging-policy.pdf> that describes the conditions of use and different charging and licensing arrangements applied for data and information. The charges levied are compliant with UK Government legislation and guidance, and links to relevant documents are included in the policy.

The BGS Intellectual Property Right (IPR) web pages <http://www.bgs.ac.uk/about/copyright/home.html> provide an advisory and licensing service for the reproduction of published material and digital map data.

This includes information on what data is available digitally, how much it will cost, terms and conditions of use, and how to apply for a license. It also gives information on copyright and commercial/ non-commercial use of data.

All recipients of a license are required to sign or digitally accept a license document detailing the terms and conditions of use before authorization is given for release of the data. A form to apply for a license is available on the BGS website <http://www.bgs.ac.uk/data/licensing/home.html>. Digital data licenses include a termination clause in case of non-compliance with the stipulated terms and conditions.

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

It should be noted that the Freedom of Information Act (FOIA) 2000, Environmental Information Regulations (EIR) 2004, Public Records Act (PRA) 1958/67 and Data Protection Act (DPA) 1998 legislation overrides any licensing agreements.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

Further information on monitoring of compliance and the specific consequences of non-compliance should be specified in order to reach compliance level 4; monitoring of license conditions but must be made explicit.



### 3. Continuity of access

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The NGDC is implemented as an integral part of the BGS infrastructure. This enables the NGDC to utilise the enterprise-level data storage (SAN), servers, systems and local/wide area networks including high-speed access to the UK Joint Academic Network (JANET).

A formal continuity plan for the NGDC does not currently exist but is under development and will be implemented in the near future. The NGDC data assets are held within and accessible from the BGS infrastructure and, as a result, their future continuity is determined by that of the organisation.

To comply with best practice and industry standards data assets are backed up and archived according to stated policies that are implemented by the BGS Systems and Networks Support (SNS) and made available internally to BGS staff via the organisational intranet. Critical data is replicated between two SAN systems in order to ensure continued access.

All data held in the NGDC is described using appropriate INSPIRE/UK government (UK GEMINI) compliant metadata [http://guidance.data.gov.uk/gemini\\_iso.html](http://guidance.data.gov.uk/gemini_iso.html). It is accessible and searchable in metadata catalogues available via the BGS/NGDC web sites and also harvested by metadata aggregator sites such as NERC Data Catalogue Service, Data.Gov.UK and EU JRC.

A digital data preservation strategy and policy for the organisation has been created that aims to encapsulate best practice from the UK and EU community (including from EU-funded FP7 projects such as the SCIENCE Data Infrastructure for Preservation (SCIDIP-ES) that developed data preservation services and toolkits for data management). Using this approach, tools and methodologies have been identified to help manage the NGDC data assets and plan activities for their stewardship to ensure future access.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 4. Confidentiality/Ethics

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The NGDC has rigorous processes in place for ingestion of data into the repository <http://www.bgs.ac.uk/services/ngdc/remit.html> including a strict quality procedure for the release of confidential data <http://www.bgs.ac.uk/services/NGDC/confidentiality.html> and guidance for data depositors <http://www.bgs.ac.uk/services/ngdc/guidelines.html>

The NGDC complies with NERC Information Security Policy and NERC Data Protection Policy to safeguard and protect the information assets in its care. (For security reasons these are internal documents that are only available to staff)

The NGDC also complies with NERC Ethics Policy <http://www.nerc.ac.uk/about/policy/policies/nerc-ethics-policy/> that provides the guiding principles applied to all aspects of the operations of NERC and its component research and data centres. It includes guidance on procedures for staff who have concerns about research procedures or identify breaches in the ethical policy. Serious concerns are referred to the NERC Ethics Board who will consider the issue and has the power to take any necessary action. The Board is accountable to the Chairman of NERC.

An internal BGS/NGDC ethics panel specifically for data access and use is in the process of being created. Its mission will be to address any issues regarding the access and re-use of the data assets held by the organisation.

NERC Research Grants and Fellowships Handbook

NERC also publishes a Research Grants and Fellowship Handbook <http://www.nerc.ac.uk/funding/application/howtoapply/forms/grantshandbook/> that includes guidance on research ethics. Researchers are required to comply with the RCUK Policy and Guidelines on Governance of Good Research Conduct <http://www.rcuk.ac.uk/documents/reviews/grc/rcukpolicyandguidelinesongovernanceofgoodresearchpracticefebruary2013-pdf/>, which should be read in conjunction with the UK Concordat to Support Research Integrity

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

<http://www.rcuk.ac.uk/funding/researchintegrity/> and the guidance on Good Research Conduct and Research Integrity <http://www.nerc.ac.uk/about/policy/policies/research-integrity/>. These policies and guidelines apply equally to researchers, support staff, research administrators, Research Council staff and all individuals contributing to the Research Councils' peer review process.

The National Geosciences Data Centre also complies with the Data Protection Act (DPA) 1998, Freedom of Information Act (FOIA) 2000 <http://www.nerc.ac.uk/about/policy/foi/information/>, Environmental Information Regulations (EIR) 2004 and Public Records Acts (PRA) 1958/1967 legislation. These legislative requirements are included in the NERC Records Management Policy <http://www.nerc.ac.uk/about/policy/foi/records-management-policy.pdf> and NERC Data Policy <http://www.nerc.ac.uk/research/sites/data/policy/data-policy/>

In cases of non-compliance with these conditions, the RCUK can invoke its' disciplinary policy to ensure that the highest standards of behaviour and conduct in research are met <http://www.rcuk.ac.uk/documents/terms/disciplinarypolicy-pdf/>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 5. Organizational infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The NGDC is funded via the BGS Informatics Directorate that in turn receives funding from NERC. NERC commissions its research centres and data centres, and provides appropriate funding to deliver the research centre, the data centre, underpinning infrastructure and the science programmes. In the future NERC is moving towards a commissioning process for the Environmental Data Centres (EDCs) that will determine the funding, services and functions it commissions from its data centres.

The co-location of the NGDC within a long-standing organisation such as the BGS ensures confidence in its ability to manage the data for the long-term, and utilises its infrastructure to deliver the appropriate services and functions. It also allows the NGDC to call upon a wide range of both informatics and geoscience domain specialists/experts for the purposes of delivering its services, much more so than would be possible if the NGDC was an entirely separate organisation.

The BGS employs over 450 scientific staff in a variety of differing disciplines who can be consulted for their input regarding the diverse range of geoscientific data encountered. The BGS Informatics Directorate is composed of 70 staff that include collections and records managers, scientific or research data managers, database or application developers and web designers as well as staff with expertise in digital preservation, scientific data accession, active data management planning, information architecture, international and regional data standards and web services. The NGDC utilises differing proportions of effort from a range of these staff in the course of delivering its functions and services.

NGDC staff have access to a comprehensive learning and development programme provided to all BGS employees, ensuring staff are up to date with new developments in IT and data management techniques through relevant training. The BGS also holds the UK Investors in People accreditation that embodies appropriate professional development strategies.

Internal governance of the data centre is through the Head of the NGDC and the Informatics Science Director that report to the BGS Executive for the purposes of delivering the agreed services and functions NERC and BGS expect from the NGDC and the Informatics Directorate.

As part of the process for continued funding of the NERC data centres a commissioning process was initiated in 2016 that included a stakeholder survey to evaluate the services each of the data centres, including the National Geoscience Data Centre (NGDC), must provide to its designated community for the future.  
<http://www.nerc.ac.uk/about/whatwedo/engage/engagement/datacentres-survey/>

The results of this survey have been used to guide the priorities and services delivered to users by the NGDC. The outcomes of this process also form the basis for planning future stakeholder engagement activities that will include user surveys, mechanisms to provide feedback on services delivered by the NGDC e.g. web-based feedback forms etc., and the reshaping of the former Information Advisory Panel (see also R6 below).

The IAP was previously an external panel composed of experts from industry and academia that provided guidance and comment on the work of the BGS Informatics directorate and the NGDC. Following the NERC data centre commissioning process, this external advisory board is now in a state of transition that will result in a revision of its Terms of Reference and review of the membership to ensure that NGDC stakeholders are fully represented.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 6. Expert guidance

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The British Geological Survey encourages and supports on-going employee training and relevant accreditation to ensure that staff, including those involved in the management and delivery of the NGDC, have appropriate and current knowledge and skills.

The NGDC staff proactively engage with a number of initiatives and organisations that provide expertise on a range of relevant topics to ensure that the services provided by the data centre are in line with the current best industry practice. This includes bodies such as the Research Data Alliance (RDA) where NGDC staff are both members and co-chairs of several interest groups and working groups e.g. Active Data Management Planning IG; metadata interest and working groups, etc., and professional associations such as the Information and Records Management Society (IRMS) <http://www.irms.org.uk/>.

A number of the staff with direct responsibility for the operation and management of the NGDC also sit on a range of relevant advisory and technical boards including:

Belmont Forum

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)



- 

Research Data Alliance Technical Advisory Board

- 

AGU Data Management Assessment Advisory Board

- 

GEO Working Groups: Data Sharing and Data Management Principles

- 

NGDC staff also regularly access the expertise of recognised organisations that provide advice and guidance on selected aspects of the data centre activities, for example:

The UK National Archives (TNA) <http://www.nationalarchives.gov.uk/> : best practice on records management, transfer, and information re-use

- 

Digital Preservation Coalition (DPC) <http://www.dpconline.org/> : guidance, good practice and tools for all aspects of creating, managing and preserving digital material

- 

Digital Curation Centre (DCC) <http://www.dcc.ac.uk/> (See R11): advice on all aspects of digital curation especially data management planning and associated tools

- 

Geoscience Information Group (GIG) <https://www.geolsoc.org.uk/gig> : affiliated group of the Geological Society that promoted best practice in use and management of geoscience information

-

NGDC staff also regularly engage in knowledge exchange with a range of organisations, including other repositories, data centres, and similar organisations around the world, e.g. Australian National Data Service (ANDS), National Centers for Environmental Information (NCEI) in the USA. This also includes the other NERC data centres (EIDC, PDC, BODC, CEDA) which interact both on an ad hoc basis and also through an internal Data Operations Group (DOG) that coordinates and advises on various aspects of data management policy across the entire research council.

As a result of the NERC survey to evaluate the services each of its data centres, including the National Geoscience Data Centre (NGDC), delivers to its designated community (<http://www.nerc.ac.uk/about/whatwedo/engage/engagement/datacentres-survey/>) the NGDC is planning future stakeholder engagement activities. These will include an online feedback form similar to that already in use for general comments and suggestions on the BGS website <http://www.bgs.ac.uk/comments/home.cfm?commentType=general>, other forms of user engagement, for example via social media, and restructuring of the previous Information Advisory Panel (IAP), an external advisory board that provided guidance and comment on the activities of BGS Informatics and the NGDC (see also R5 above).

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 7. Data integrity and authenticity

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The NGDC has a data ingestion policy in place that sets out the requirements of depositing data with the data centre. <http://www.bgs.ac.uk/services/ngdc/remit.html> The requirements include UK Government GEMINI standard compliant metadata, terms, and conditions.

The metadata must be complete before the data can be ingested. A NGDC Data Deposit Portal has been developed for receiving deposits of data, including the required metadata, by direct upload. <http://transfer.bgs.ac.uk/ingestion> It is the responsibility of the Data Ingestion Team to ensure that the data and the metadata is complete before finalising its ingestion into the NGDC. Where the deposit is incomplete, the Ingestion Team will liaise with the depositor. If data does not meet the data value checklist, it is returned to the depositor.

The NGDC uses an ORACLE® database system to store metadata and links to the digital objects that are deposited. Internal persistent identifiers are used to track a deposit and ensure internal system integrity, and also feed into NGDC data citation procedures. Where requested, the repository can mint a digital object identifier (DOI) using the DataCite service provided by the British Library.

Automatic fixity checks on deposited data are not in place at present, but their use is being investigated. The original deposited digital object is kept in a separate store to the processed digital object. The completeness of the data and associated metadata is captured as part of the metadata record. This is stored in an ORACLE® database that includes history tables to track any revisions.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

In future assessments the archive may wish to clarify its PID policy to explain the nature of the “internal persistent identifiers” and the policy which appears to make assignment of DataCite identifiers a choice of the data producer.

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

## 8. Appraisal

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

Data deposits are appraised against the Data Collection Policy and the Data Value Checklist (<http://www.bgs.ac.uk/services/ngdc/documents/DVCNGDC.pdf>). Following the Data Ingestion workflow, the data is checked for completeness. The NGDC has a list of preferred data formats (<http://www.bgs.ac.uk/services/NGDC/preferredDigitalFormats.html>) and any data supplied in other formats requires dialogue with the donor. The data can be rejected and sent back to the depositor during this process, or further information requested from the depositor.

A NGDC Data Deposit Portal (<http://transfer.bgs.ac.uk/ingestion>) has been developed for depositors to upload data and provide basic metadata to describe the deposit (as described under R7), including the specific terms and conditions of discovery and re-use for the data. The metadata required depends on the type of deposit, but all generic deposits are based on the INSPIRE/UK government (UK GEMINI) Standard. All depositors are encouraged to use the Portal, however, if a deposit is too large for upload alternative options are available for submitting the data and associated deposit form.

A Deposit Form is required for all deposits, and data will not be accepted unless this is provided. It records a minimal set of metadata that allows for discoverability and re-use of the data using the UK GEMINI standard. The NGDC also creates Discovery Metadata (<http://www.bgs.ac.uk/discoverymetadata/>) for the datasets it holds that complies with the ISO standard 19115:2003 for geographic information metadata <https://www.iso.org/standard/26020.html>. A policy is currently being drafted for legacy collections where GEMINI compliant metadata may not be available.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 9. Documented storage procedures

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The NGDC undertakes data storage according to documented processes and procedures in line with NERC Information Security Policy and NERC Information Security Incident Response Procedure (internal document: available on request).

The NGDC provides guidance on the transfer of data to the data centre via the BGS website <http://www.bgs.ac.uk/services/ngdc/guidelines.html>. The original data (Submission Information Packages (SIP)) is virus checked before being stored on the network. A copy of the data is accessioned (indexed) and stored in the Donated Data Store where unvalidated data (Archival information packages (AIP)) are stored. Priority data are entered into relevant collections/ databases/ websites (Disseminated Information Packages (DIP)). Data may be normalised into other formats where required (e.g. conversion from PDF to TIFF). The Accession and Ingestion Team deal with the data and custody transfer, notifications and queries.

All data is stored on the BGS Storage Area Networks (RAID 5 compliant) at the Edinburgh or Keyworth offices and presented as Windows® file shares in Active Directory. External access to the SAN and file shares is blocked and controlled with Check Point Firewalls. Authorisation of access to the data is implemented using Active Directory file permissions that only allow authorised users to have access to the data, either read-only or read-write as appropriate. All data is backed-up daily using IBM Tivoli Storage Manager (further details of the security procedures implemented within the data Centre are described in Section R16).

The BGS Risk Register documents significant corporate risks using a scoring system, and outlines appropriate mitigation scenarios. It is reviewed and updated annually by the BGS Business Assurance Manager to ensure it remains current and fit for purpose.

Disaster recovery procedures are in place that include data recovery provisions which involve restoring (or retrieving in the case of archives) data from tapes using IBM Tivoli Storage Manager (TSM).

Daily automated scripts run to ensure that any changed data is incrementally copied to other sites. The repository follows the best practice guidance as described in the preservation policy currently under development. This includes the ISO14721 (OAIS model) for storage as well as the other preservation functions. Checksum procedures will also be initiated in the near future.

NGDC/BGS use Linear Tape-Open (LTO) to store backups and archives. The data was migrated from LTO3 to LTO4 in 2013 and will be upgraded to LTO6 in the next financial year (2107/2018). Migrating to newer LTO versions helps to ensure against media deterioration. The LTO media is stored in fire suppressant, secure rooms and LTO4 has an expected durability of 11,200 end-to-end passes, which is a figure that BGS systems are unlikely to ever approach. On the rare occasion when TSM reports a tape error, the data is migrated to another tape and the tape with the error is securely destroyed.

Daily logs are produced by the TSM servers, which alert administrators of any errors or warnings. Logs and alerts are also generated by the SANs regarding failed disks, storage capacity warnings and other hardware and software issues. These logs are emailed to several members of the systems team for immediate action

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 10. Preservation plan

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

3. In progress: We are in the implementation phase.

### *Self-assessment statement:*

The NGDC's role as the national geoscience data centre assumes an indefinite retention of the data in its care. The present RCUK data policy states that all research data should be retained for 10 years after it was last used.

The NGDC has a new digital preservation strategy and policy that will be published shortly (currently an internal document that is available on request) and is also in the process of preparing an accompanying work plan. This will include a digital preservation capability, maturity and risk assessments, and takes into account the risk of technology and file format obsolescence as well as the skills and other resources required to develop and maintain a preservation programme.

The NGDC has a robust discovery metadata schema based on the ISO 19115 standard (<http://www.bgs.ac.uk/discoverymetadata/>), and there are plans to add a preservation metadata component to key datasets as extensions to this schema. This extension will be based on the Library of Congress' PREMIS (<http://www.loc.gov/standards/premis/>) data dictionary. The NGDC is also looking into employing a checksum system to monitor against unplanned changes within its data assets not stored in a database system.

The NGDC currently migrates data formats as and when required but does not currently have a regular migration schedule in place. When data contained in the corporate ORACLE® database is migrated to a new version of ORACLE® the previous version of the system will be archived including the corporate work tasks and integrity checks used to migrate the entire data store to the new version. Both the original formats/bitstreams and any converted data are preserved.

Data ingested through the online digital NGDC Data Deposit Portal into the corporate Digital Accessions Database receives an internal persistent identifier and depositors also have the opportunity to request a DOI for their datasets. All data are assigned unique and persistent identifiers at the point of ingestion into the repository, and these identifiers will persist for the lifetime of the original deposit as part of an archive. Accessioned items for the deposit are also assigned an internal unique identifier, which persists for the life of that Accession. In addition, all items e.g. boreholes/images are assigned a persistent identifier, which exists for the life time of that entity.



The NGDC web pages provide data depositors with the information and forms necessary to lodge their digital data with the data centre <http://www.bgs.ac.uk/services/ngdc/guidelines.html>. The documentation available to the depositor includes a description of the data collection remit of the NGDC, a deposit form capturing the terms and conditions for access and long-term storage of the data, a metadata form, a list of preferred formats, and a data value check list. The ingestion policy and the collection remit define the mandate of the repository with regard to long-term storage of the data, and the tools enable the capture of all relevant information and support the longevity of the data during and after the transfer to the NGDC.

The NGDC has a well-established and supported data management planning procedure in place for both internal data (on the corporate intranet) and NERC grant data that takes into account the whole data lifecycle (<http://www.nerc.ac.uk/research/sites/data/dmp/>). The repository staff monitor developments within the field of digital preservation and collaborate with other preservation and curation organisations such as the Digital Preservation Coalition (DPC) and the Digital Curation Centre (DCC) in the UK.

BGS has also participated in a number of relevant current and previous EU-funded projects, such as SCIENCE Data Infrastructure for Preservation (SCIDIP-ES), with the aim of expanding and developing its' knowledge and expertise in the field of digital data preservation. The benefits derived from involvement in this type of activity include the innovative tools and methodologies that can be implemented within the NGDC to improve the effectiveness of its data management processes.

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

See previous comments regarding need to provide further information on PID policies and practices. We also note that this response clarifies lack of preservation mandate proof in criterion 1, but the information needs to be provided there as well.

## 11. Data quality

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

The NGDC Data Policy requires that all datasets have comprehensive discovery metadata. The NGDC Data Deposit Portal used to receive data, metadata, and to document the terms of deposit (i.e. restrictions imposed on the re-use/dissemination of the data by the depositor), employs the UK GEMINI metadata standard for all generic deposits. Specific deposits relating to particular areas such as NERC Grants, and BGS internal deposits also use ISO 19115:2003 (Geographic Information Metadata) and INSPIRE standards. The repository uses an ORACLE® database to store metadata and maintains an audit trail for all content.

Discovery metadata for the datasets held by the NGDC is made available on the BGS website: <http://www.bgs.ac.uk/discoverymetadata/>. This discovery level metadata is also harvested by a number of systems including Data.gov.uk and the NERC Data Catalogue Service. The catalogue of data holdings is also made available as an Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW) <http://www.bgs.ac.uk/data/services/csw.html>.

Technical adherence to metadata standards and overall quality of the metadata is also checked. The discovery metadata is also scrutinized by other peers and users through an annual review conducted by the metadata subgroup of the NERC Data Operations Group (see section R6).

NERC expects the scientific quality of the data generated by its funded grants or programmes to be ensured and maintained by the scientific staff working on the funded grant or programme. It is not feasible for the NERC Environmental Data Centres (EDC), one of which is the NGDC, to check the scientific quality of all the diverse geoscience data it receives. The EDC will however ensure that the data supplied has appropriate metadata to describe the data asset and data files have headers, units of measurement, consistent population and other quality checks.

### Reviewer Entry

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

*Accept or send back to applicant for modification:*

Accept

*Comments:*

In future submissions please also include feedback from users.

## 12. Workflows

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The NGDC has policies and procedures in place to cover the lifecycle of the data from the pre-ingestion phase to digital preservation.

The ingestion begins with the data collection policy <http://www.bgs.ac.uk/services/ngdc/remit.html> and the data value checklist <http://www.bgs.ac.uk/services/ngdc/documents/DVCNGDC.pdf>. All NGDC data acquisitions are in line with the policy and the checklist. The process is implemented using a data ingestion workflow, which covers the detailed accessioning of data, data prioritisation and processing.

The ingestion procedures for the repository staff include a strict quality procedure for the release of restricted data <http://www.bgs.ac.uk/services/NGDC/confidentiality.html>.

The NGDC provides a Data Deposit Portal for upload of data to the repository <http://transfer.bgs.ac.uk/ingestion>. There are also alternative methods available for submitting larger datasets that are agreed with the depositor. Data depositors are guided through the deposit process via the NGDC webpages <http://www.bgs.ac.uk/services/ngdc/guidelines.html>. They are provided with online deposit forms that must be completed for all submissions before data is ingested into the repository. These forms capture information relating to the terms and conditions for data access and long-term storage as defined by the depositor, metadata, etc.

Data ingested through the online digital NGDC Data Deposit Portal into the corporate Detailed Accessions Database (<http://www.bgs.ac.uk/services/NGDC/dataDeposited.html>) receives an internal unique identifier, and depositors are also offered the possibility of obtaining a DOI for their datasets on request <http://www.bgs.ac.uk/services/NGDC/citedData/catalogue.html>. All data is assigned an identifier at the point of entry to the repository, and these internal identifiers will persist for the lifetime of the original deposit as part of an archive.

The NGDC undertakes data storage against documented processes and includes procedures in line with NERC Information Security Policy and NERC Information Security Incident Response Procedure (internal documents

available on request) The repository complies with DPA 1998, FOIA 2000 <http://www.nerc.ac.uk/about/policy/foi/information/>, EIR 2004 and PRA 1958/1967 legislation. These are included in the NERC Records Management Policy

<http://www.nerc.ac.uk/about/policy/foi/records-management-policy/> and NERC Data Policy <http://www.nerc.ac.uk/research/sites/data/policy/data-policy.pdf>

The NGDC has a new digital preservation strategy and policy that it is in the process of being finalised and formally published along with an accompanying work plan (as described under section R10) which will include appropriate workflows for digital preservation. Public access to the collections policy, the data value checklist, the preservation policy and the Open Government License terms and conditions applicable to many of the data, ensures transparency throughout the data selection and archiving process.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### 13. Data discovery and identification

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### Applicant Entry

*Statement of Compliance:*

3. In progress: We are in the implementation phase.

*Self-assessment statement:*

NGDC provides direct access to a range of key datasets through its OpenGeoscience service that also allows users to view maps, images and information. It also supports discovery of the data it holds through the BGS Discovery Metadata service <http://www.bgs.ac.uk/discoveryMetadata/home.html>. The user interface provides functionality to interrogate the NGDC data catalogues using various search criteria including keywords, geographical location etc. Users can also directly browse alphabetical lists of the available datasets and access detailed descriptions of the individual datasets.

The NGDC also has a documented data citation process for selected datasets that is available on its website <http://www.bgs.ac.uk/services/ngdc/citedData/home.html?> Information is provided on the rationale that is employed for selecting these 'approved' datasets and the associated data citation process. The NGDC provides a number of datasets that have undergone a rigorous process to ensure their validity and integrity before being assigned a DOI and included in the associated data catalogue.

NGDC is also an issuing agent for DataCite DOIs that allows direct citation of the datasets that it holds. In order for the NGDC to issue a DOI for a dataset it must be fully ingested into the datacentre to ensure that it is of the required quality, has all of the necessary supporting information and available for re-use following citation.

#### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

The user experience would be improved via DOIs that are directly available with each download.

## 14. Data reuse

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Discovery metadata complying with INSPIRE 19115/19139, which in the UK also complies with the UK GEMINI v2.2 schema, is required when data is received by the repository. This discovery metadata ensures datasets are described in sufficient detail to be found using search parameters that include geographical coordinates or location, free text against title or abstracts, keywords, formats etc. This also enables datasets to be exposed through appropriate external gateways (e.g. data.gov.uk, and the NERC data catalogue).

The NGDC provides a list of preferred file formats for deposit at <http://www.bgs.ac.uk/services/ngdc/preferredDigitalFormats.html> to encourage deposit of data in formats that are at less risk from technology/software obsolescence, and which provide efficient migration paths to newer file formats when necessary to ensure digital continuity. These preferred formats are those commonly used in the NGDC's designated community and include PDF, Microsoft® Office formats for documents e.g. doc/docx, xls etc. as well as generic formats e.g. CSV, TXT, and in some instances other database formats for raw data. Spatial data is typically submitted in ESRI ArcGIS formats.

The development of a digital preservation policy and subsequent strategies to ensure the continued usability of the data have been a key element of the NGDC data management procedures. These policies and strategies include identification of migration pathways for file formats that may be under threat from technology/software changes to ensure that the data can be migrated to more future-proof formats. The aim of the NGDC is to prioritize the migration of legacy data formats to current formats to ensure the continued usability of datasets. A typical example would be data, which has been provided in older spreadsheet formats (e.g. earlier versions of Microsoft Excel, or obsolete Lotus formats) which has been converted into CSV, TXT, or ASCII formats.

In order to ensure continued understandability of the data, appropriate contextual metadata (as well as discovery metadata) is captured at the data ingestion stage. The data submitted may include readme files or higher-level guidance (contextual metadata) to ensure that it can be more easily re-used. Guidance and checking by the data centre at the point of accession and ingestion ensures that data files or spreadsheets include header rows, units and consistent population of fields.

The NGDC is also frequently required to store a range of datasets generated by Earth Science and other environmental models. In order to encourage the re-use of this data a rich level of contextual metadata is captured

using the metadata schema created as part of the NERC funded PURE (Probability Uncertainty and Risk in the Environment) initiative <http://model-search.nerc.ac.uk/>. Models are supported by appropriate metadata that provide pointers to the versioned code repository, input or output data, and papers or further guidance notes.

The NGDC recognizes the importance of providing as much contextual information as possible for its data holdings in order to encourage its re-use. The aim is to capture as much supporting information as possible (e.g. reports, manuals, and references to peer-reviewed publications) alongside the raw data, so that users of the repository can access these resources when re-using the data. By providing this additional supporting information the NGDC ensures that users can make an informed assessment of whether individual datasets are fit for their specific purpose and therefore make appropriate use of it.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 15. Technical infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

#### **Storage / Infrastructure**

Online storage for the data centre repository is provided by Hitachi and Dell Storage Area Networks (SAN) administered by a highly skilled in-house IT infrastructure team. Data that does not require on or near-line storage is archived to Enterprise Tape Libraries, implemented using IBM Tivoli Storage Manager (TSM), and supported by a comprehensive archive metadata system.

The SANs and Enterprise Tape Libraries provide a secure storage environment with well-maintained back-up and maintenance routines (backup and security procedures are explained more fully in section R16.)

The rate of accumulation of stored data is closely monitored so that requirements for increases in data storage capacity can be planned in advance in the light of evolving requirements. The SAN infrastructure described above is designed to be readily extensible (by, for example, the addition of integrated expansion modules).

The BGS Software Licence Manager maintains a software inventory, including documentation covering the local IT infrastructure, that is maintained internally.

The BGS technical infrastructure is based primarily on proprietary software (e.g. IBM Tivoli Storage Manager to implement the enterprise tape library, and ORACLE® databases). Where community supported software is in use, formal maintenance agreements are in place. Examples include CentOS and Ubuntu Server. Users access the BGS technical infrastructure via the UK Joint Academic Network (JANET), over dual-redundant 1Gbps links providing 24 hour a day access to NGDC data and maximizing service availability.

#### **Ingestion**

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

The technical infrastructure of the NGDC is built upon appropriate international and relevant domain standards to ensure that it is scalable, extensible and readily maintainable. For example, the discovery metadata captured as part of the data ingestion process conforms to INSPIRE 19115/19139 metadata standards, which also complies with the UK government GEMINI v2.2 schema. The ISO19115 and 19139 standards are extensive and the data centre therefore implements a core schema of the key metadata elements relevant to the data provided by the designated NGDC communities.

## **Delivery**

Delivery of smaller data packages can be via direct download from the NGDC data portal. The NGDC also provides access to repository data via OGC compliant web services made available through the British Geological Survey website, for example:

- British onshore digital geological maps (1:50,000): <http://www.bgs.ac.uk/data/services/digmap50wms.html> .
- UK onshore bedrock geological data at 1:625 000 scale WFS services <http://www.bgs.ac.uk/data/services/wfs.html>

Using OGC standards such as Web Feature Service (WFS) and Web Map Service (WMS) enables the NGDC to provide data in consistent formats for consumption by the designated user communities who routinely utilise these standards (e.g. for accessing data in a map-based environment).

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

Accept

*Comments:*

## 16. Security

### *Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

### *Statement of Compliance:*

3. In progress: We are in the implementation phase.

### *Self-assessment statement:*

The NGDC repository has a comprehensive suite of procedures in place to ensure rapid recovery and return to normal operations in the event of a system failure/disaster or other technical failure.

All data is stored on the BGS Storage Area Networks at Edinburgh or Keyworth and delivered as MS Windows file shares in Active Directory. External access to the SAN and file shares is blocked and controlled with Check Point Firewalls. Published data is in a read-only format. External access to the unpublished data is only available via a Secured VPN with multi-Factor Authentication using Digital Tokens. Carbon Black Protection is running on the SAN nodes as well as all End Points to block any malicious activity that may corrupt or damage the data. Access to the data is authorised using Active Directory file permissions that allow only privileged user's access to the data, either read-only or read-write as appropriate.

All data is backed-up daily using IBM Tivoli Storage Manager. The data will be part of a retention and recovery schedule that allows a rolling three months' worth of file retrieval. A copy of the tape archive is securely stored off site in-line with the NGDC's disaster recovery and archiving policies. There are replicas of data copied to both sites (Edinburgh and Keyworth) and communication between the sites is secured using IPSEC protocols.

The NGDC utilises the expertise and skills of the full time BGS Information Security Officer (ISO). The ISO is responsible all aspects of implementing and maintaining the security of the BGS and NGDC IT infrastructure.

### **CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

Functionally the ISO provides technical support for all aspects of BGS/NGDC cybersecurity that includes configuration of the firewalls, and provision of data access for external users via the secured VPN and active directory structures. BGS has received the Cyber Essentials Plus accreditation (<https://www.cyberaware.gov.uk/cyberessentials/>), and attained the ISO:9001:2008 Quality Management Standard for its quality management systems.

The system of back-up procedures and storage of multiple copies of data at geographically separate sites, described above, forms a key component of the disaster recovery and business continuity procedures, providing for rapid recovery of data and infrastructure under commonly anticipated threats (e.g. technical failure, human error). This system also ensures the safety of the data in the event of a more serious incident where, for example, the buildings housing the data centre and/or major IT infrastructure were to be rendered inoperable.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 17. Comments/feedback

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The British Geological Survey and/or NGDC hold a number of ISO, staff and environmental accreditations supporting the organisation and demonstrating its capability in the areas of management systems, commitment to staff, IT/IS security and fair trading. <http://www.bgs.ac.uk/about/accred.html>

The NGDC staff involved in the DSA repository certification process have found it to be clear, robust and systematic in its approach. The DSA process has helped the NGDC to scrutinise and review existing systems, processes and policies. It has also encouraged the organisation to benchmark its' procedures against the questions and criteria supplied as well as consider further enhancements and improvements to current workflows and best practices. As a result of the NGDC undertaking this process other NERC data centres are now considering also undertaking the DSA process.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*