



National Geoscience Data Centre

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

-

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

National repository system; including governmental

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

Brief Description of Repository

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) as part of its Digital science area. The NGDC/BGS is funded via its parent organisation the Natural Environment Research Council (NERC).

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.

The National Geoscience Data Centre is also one of five federated data centres that form the UKRI-NERC Environmental Data Service (EDS) which provide data centre functions and services across the range of scientific communities funded within the research council <https://nerc.ukri.org/research/sites/environmental-data-service-eds/>. NGDC delivers the geoscience/sub-surface element of this data service, which covers the thematic and regional data created by projects and research funded by NERC as one of the component research councils within UKRI.

The Environmental Data Service is approaching the end of year 4 (April 2021) of an initial 5-year commissioning period, and the process of reviewing the EDS outputs and impacts has recently commenced as a precursor to defining the specifications for the NGDC functions/activities as part of the next commissioning cycle that will begin shortly. It is hoped that the next commissioned cycle of the EDS will have a 10-year duration to facilitate the evolution, development and enhancement of the services and functions across the five data centres, and support better utilisation of UKRI-NERC infrastructure, shared facilities and best practices. This clarity and longer commissioning cycle will place the NGDC in a stronger and more secure funding position with respect to longer-term professional data management.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

Brief Description of the Repository's 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. A diagram of the context and landscape of the data centre is included with this application.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept.

Note: Apparently, the content diagram has been attached to an email by NGDC, see NGDC_feedback.pdf). Yet, apparently, it has been not linked to the document itself.

Level of Curation Performed. Select all relevant types from:

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

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

Comments

The goal of the NGDC is the long-term and professional archiving, preservation and dissemination of its data holdings. This data forms an evidence base for previous and current research programmes, underpins existing scientific products, and a critically important scientific resource for inclusion in future science projects, programmes, applications, systems, products or decision support systems. The NGDC facilitates open data access and re-use including adoption and implementation of the FAIR principles that also supports robust scientific peer-review processes and the generation of wider impact for UK industry and businesses.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

Insource/Outsource Partners. If applicable, please list them.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

n/a

Summary of Significant Changes Since Last Application (if applicable).

NGDC Data Centre: restructuring

BGS undertook a refresh of its science strategy during 2017-2018 which also included a reorganisation to enable it to better deliver new strategic priorities and projects. <https://bit.ly/3pRA4Hi>. Part of that reorganisation was a recognition that 'digital' functions needed to be represented at a higher level within the organisation and be seen alongside the geoscience domains (three primary geoscience challenges and one digital). The post of Chief Digital Officer (CDO) was also created with a seat on both the Senior Strategy Group and Senior Management Board. The CDO was appointed in 2019 with a mandate to reorganise the digital area to better reflect the new Digital Strategy (<https://bit.ly/2L5Mj4f>).

Four digital areas were created as part of the implementation of the new digital strategy that include: Data Centre, Digital Labs, Products, Idealisation and Enterprise, and IT Security, Infrastructure and Apps Development. The vision for the Data Centre was to manage the entirety of the organisations (BGS and NGDC) digital and analogue data be it scientific, business and/or operational, and in accordance with best practices within a robust data governance framework.

This restructuring exercise has also provided the NGDC with an opportunity to review current operations and conduct an information gathering exercise among its broad range of stakeholders that has led to the development of a core strategy for the NGDC going forward. This strategy includes a number of key actions to maintain and improve its trusted user services which includes retaining its Core Trust Seal repository certification. Other key actions include implementation and adoption of new technologies, best practices and protocols to optimise data management processes and maximise user engagement both within and across domains. The overall aim of this entire restructuring exercise is to further improve and advance the current operations and user services delivered by the NGDC, many of which are aligned with recommendations made as part of the previous CTS evaluation process.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

Other Relevant Information.

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

BGS has been in existence since 1835, under a number of different organisational names, and is one of the oldest geological surveys in the world. 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 held by NGDC 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
<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/>

The NGDC has a range of policies that underpin the functions and services it delivers, including the NERC Data Policy, BGS Research Data Management Policy, metadata, collections, data management planning, digital preservation and acceptable file formats. Further details of the relevant NERC data policies, guidance etc. can be accessed at:
<https://nerc.ukri.org/research/sites/environmental-data-service-eds/policy/>

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

ORGANIZATIONAL INFRASTRUCTURE

1. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository
Accept.

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
Accept

Response:

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 (<https://nerc.ukri.org/research/sites/data/policy/>) Ingestion Policy (<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/>) and Digital Preservation Policy (<https://www.bgs.ac.uk/download/bgs-digital-preservation-policy/>) encapsulate the value of scientific data and the importance attached to its long-term professional management and preservation as both an evidence base for existing scientific projects and 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 (le), Science and Technology Act 1965 and Water Resources Act 1991 (198 & 205).)
- Scientific programme outputs from the national geological survey (both in the UK and globally)
- 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 UKRI data policy states that all research data should be retained at the minimum for 10 years after it was last used. The NGDC has introduced a preservation programme to manage long-term digital preservation and to assume a stewardship role to ensure the survival of digital materials from one generation of digital systems to the next. In an article based on a MSc dissertation in 2016 (<http://nora.nerc.ac.uk/id/eprint/517250/>) one of the key findings was that the validity of geoscience data held at the NGDC is much longer than 10 years, and where the work was funded by the UK Government the data may be held in perpetuity under the Public Records Act. These two points take precedence over the UKRI 10-year retention policy.

The NGDC is embedded within the Digital research area of the BGS, with the aim of efficiently aligning 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 <https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/>

NGDC also delivers the geoscience/sub-surface component of the UKRI-NERC Environmental Data Service (EDS) that includes the thematic and regional data created by all projects and research funded by NERC, which is one of the component research councils within UK Research and Innovation (UKRI) that is the responsible funding agency.

Initially commissioned in 2017 for a 5-year period, the EDS is delivered by a federated group of five data centres including NGDC. As this first period is now coming to an end, a new commissioning process will commence in the near future that will include reviewing the EDS outputs and impacts, and defining the specifications (Data Centre functions/activities) for the next phase. It is hoped that the next commissioned cycle will be for 10 years to provide structured longer-term investment to support the ongoing development of the EDS and enhance the service and functions across the five data centres including better utilisation of UKRI-NERC infrastructure, shared facilities and best practice.

The EDS is underpinned by a large-scale data management framework and digital infrastructure that includes both centralised high-performance computing services, delivered by JASMIN <https://www.jasmin.ac.uk/> , and local computing infrastructure, which ensures the longer-term sustainability of data and other digital research assets.

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:

Accept

2. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Accept.

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Accept

Response:

Repository comment on 1st February 2022: "Reviewers are presently unable to agree on a scoring - we originally marked ourselves as '4' fully implemented in the repository but in a previous review, one reviewer marked us down to '3' but currently the reviewers do not agree (one '3' and one '4'). As there are challenges in the UK landscape on licensing for OGL, CC BY4.0 and agreed with '3' implementation phase marking. We are happy with either '3' or '4' but unable to ensure two anonymous reviewers agree. In the October 2021 review both reviewers accepted '3' rating. We didn't expect to need to re-do/supply more information for this criteria when it was previously agreed and there were no reviewers comment to address."

To encourage the use and re-use of data the NGDC makes data freely available wherever possible via the BGS website at <https://www.bgs.ac.uk/geological-data/opengeoscience/?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.

The NGDC complies with NERC Data Licensing and Charging Policy <https://nerc.ukri.org/research/sites/data/policy/nerc-licensing-charging-policy/> 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 data licensing web pages <https://www.bgs.ac.uk/information-hub/licensing/> provide an advisory and licensing service for BGS datasets. 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. All recipients of a BGS license are required to sign a license document detailing the terms and conditions of use before authorization is given for release of the data. Digital data licenses include a termination clause in case of non-compliance with the stipulated terms and conditions.

It should be noted that the provisions of the Freedom of Information Act (FOIA) 2000, Environmental Information Regulations (EIR) 2004, Public Records Act (PRA) 1958/67 and UK Data Protection Act (DPA) 2018 legislation with regard to the duty to disclose data override the confidentiality provisions of any individual licensing agreements, although those licensing provisions will be taken into account. If BGS already makes specific information available under licence then requests made for the same information under this legislation may be dealt with by reference to the existing licence arrangements available to the enquirer. However, this policy is currently under review as a result of the work done as part of the CTS self-evaluation process.

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).

BGS also lead and participate in UK Geospatial Commission work on machine readable data licensing that will be used in the future for online accessible resources. <https://www.gov.uk/government/organisations/geospatial-commission>

The BGS Intellectual Property Right (IPR) web pages (<https://www.bgs.ac.uk/bgs-intellectual-property-rights/>) provide an advisory and licensing service for the reproduction of published material and digital map data. The NGDC have direct access to BGS experts in this area.

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. Digital data licenses include a termination clause in case of non-compliance with the stipulated terms and conditions.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

3. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Acceptable.

Reviewer 2

Comments:
3 – The repository is in the implementation phase
Accept

Response:

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. BGS is an approved Place of Deposit under the Public Records Act 1958 and maintain the archives on behalf of the UK National Archives.

<https://www.nationalarchives.gov.uk/archives-sector/legislation/approved-places-of-deposit/>

<https://www.legislation.gov.uk/ukpga/Eliz2/6-7/51>. If BGS were unable to continue in this role for some reason then the responsibility for public records would transfer back to The National Archives (TNA).

TNA are the designated national authority for data archiving in the UK, and they hold considerable amounts of data on behalf of the nation. They also understand the specific requirements for specialist knowledge, long-term storage, management and care of different data types. The National Archives (TNA) formally identifies the NGDC/British Geological Survey as a 'Place of Deposit' for geoscience data in the UK, recognising the volume and significance of our geoscience data holdings, and associated domain level expertise.

The relationship between the TNA and NGDC also ensures that we are aligned with their best practices, guidelines and recommendations, which includes a periodic inspection to check compliance. In addition, as a designated Place of Deposit, if our funding or support was deemed insufficient or terminated for some reason then TNA would assume responsibility for our geoscience data holdings. However, any plan for this eventuality is likely to be highly complex (and potentially time limited), and would therefore only be developed if it became clear that this transfer of responsibility for our data holdings to TNA was necessary.

In addition, both organisations recognise that this situation is highly unlikely to materialise when the NGDC geoscience data sets are core to BGS functions for both scientific research and national public good roles including maintaining and delivering this data for all external users. As the NGDC is hosted by the BGS, a situation that requires TNA intervention could only arise if BGS itself became non-viable for some reason. However, BGS is the national geological survey that has existed in some form for more than a hundred years, and the future of the organisation is likely to be secure for many years to come.

To comply with best practice and industry standards the NGDC 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 https://guidance.data.gov.uk/publish_and_manage_data/harvest_or_add_data/harvest_data/gemini/. 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.

In an effort to drive digital transformation across the organisation NGDC is also initiating a digitalisation programme to maximise access to our national collections as described in the BGS digital strategy <https://www.bgs.ac.uk/download/bgs-digital-strategy-2020-2025>. A digital data preservation strategy and policy for the organisation aims to encapsulate best practice from the UK and EU communities. They form a modular preservation programme or framework for our preservation activities which are updated annually as resources and requirements change. NGDC has carried out a digital research data survey amongst BGS researchers to canvas their research data management (RDM) practices and requirements. As a result, the BGS Research Data Management Policy [https://www.bgs.ac.uk/download/bgs-research-data-management-policy/? undefined=undefined](https://www.bgs.ac.uk/download/bgs-research-data-management-policy/?undefined=undefined) and an internal report on the findings (available on request) have been produced to support planning of the NGDC service delivery and digital preservation programme. The findings of the report have also fed into redesigning our existing RDM training course https://ipres2019.org/static/pdf/iPres2019_paper_12.pdf, which was originally aimed at PhD students but will now also cater for researchers and external stakeholders. These training materials will be modified so they can also be used as part of our online guidance and made available as webinars for staff and students. The cornerstone of our approach is enabling data creators to follow best practice leading to robust data deposits being transferred to the repository.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

4. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository
Accept.

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
Accept

Response:

The NGDC has rigorous processes in place for ingestion of data into the repository

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/> including a strict quality procedure for the release of confidential data <https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/#nerc> and guidance for data depositors.

The NGDC complies with UKRI/NERC Information Security Policy and UKRI Data Protection Policy

<https://www.ukri.org/files/termsconditions/ukri-data-protection-policy-pdf/> to safeguard and protect the information assets in its care. (For security reasons the former is an internal document that is only available to staff)

BGS is committed to good practices related to research ethics and integrity to ensure the dignity, welfare and rights of all those involved are respected (whether they are participants, third parties or BGS staff). A BGS Research Ethics Committee has recently been established that includes a member of the NGDC staff. In addition, BGS has developed its own Research Ethics and Research Integrity Policy

<https://www.bgs.ac.uk/download/bgs-research-ethics-and-integrity-policy-2/> which provides a general framework for its research practices and also promotes higher ethical standards in geoscience.

This policy also takes into consideration and, where possible, aligns with relevant legislation. For example, the policy states that:

“Data and records storage and management practice should align with the UKRI Data Protection Policy, UKRI and NERC Data Policy <https://nerc.ukri.org/research/sites/data/policy/> and UKRI funding policy <https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>, and where appropriate UKRI Records Management Policy <https://bbsrc.ukri.org/documents/ukri-records-management-policy-pdf/> with its associated retention schedule. The processing (acquiring, holding, using, etc.) of personal data in the UK is governed by the EU General Data Protection Regulation (GDPR) and the UK Data Protection Act 2018.”

The NGDC also complies with NERC Ethics Policy <https://nerc.ukri.org/about/policy/policies/nerc-ethics-policy-qs/> 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.

NERC Research Grants and Fellowships Handbook

NERC also publishes a Research Grants and Fellowship Handbook <https://nerc.ukri.org/funding/application/howtoapply/forms/grantshandbook/> that includes guidance on research ethics.

Researchers are required to comply with the UKRI/RCUK Policy and Guidelines on Governance of Good Research Conduct <https://www.ukri.org/files/legacy/reviews/grc/rcuk-grp-policy-and-guidelines-updated-apr-17-2-pdf/> , which should be read in conjunction with The Concordat to Support Research Integrity <https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2019/the-concordat-to-support-research-integrity.pdf>. These policies and guidelines apply equally to researchers, support staff, research administrators, UKRI Research Council staff and all individuals contributing to the Research Councils' peer review process.

The National Geoscience Data Centre also complies with the Data Protection Act (DPA) 2018, 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 UKRI Records Management Policy <https://bbsrc.ukri.org/documents/ukri-records-management-policy-pdf/> and NERC Data Policy <https://nerc.ukri.org/research/sites/data/policy/>

In cases of non-compliance with these conditions, the UKRI can invoke its code of conduct to ensure that the highest standards of behaviour and conduct in research are met

<https://www.ukri.org/our-work/supporting-healthy-research-and-innovation-culture/research-integrity/>

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

5. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase
Accept.

Reviewer 2

Comments:

3 – The repository is in the implementation phase
Accept

Response:

The NGDC is funded via the BGS Digital Challenge area that in turn receives funding from UKRI-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. The National Geoscience Data Centre (NGDC) delivers the geoscience/sub-surface element of the UKRI-NERC Environmental Data Service (EDS) which is comprised of five federated data centres that serve a range of scientific communities funded within the research council.

<https://nerc.ukri.org/research/sites/environmental-data-service-eds/> .

The EDS was initially commissioned for a 5-year period beginning in 2017, and a new commissioning cycle for the next phase will begin during 2021. This process will start with a review of the outputs and benefits delivered by the EDS, which will lead to the re-shaping of the specifications for the data centre functions/activities for the next phase. It is hoped that the next commissioned cycle of the EDS will be for 10 years to allow longer-term planning and investment to support further development and enhancement of services and functions across the five data centres, and also ensure better utilisation of UKRI-NERC infrastructure, shared facilities and best practice. This longer commissioning cycle for the EDS will provide the NGDC with more secure funding and place the data centre on a stronger footing with regard to long-term professional data stewardship.

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 from a variety of different disciplines who can be consulted for their input regarding the diverse range of geoscientific data held by the NGDC. The BGS Digital Challenge area is composed of 70 staff that include collections and records managers, scientific or research data managers, data scientists, 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 expertise 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 they are kept up to date with new developments in IT and data management techniques through relevant training. Staff training in data management that is based on that already offered to external stakeholders, is currently under development to ensure consistent quality standards. Modules will be developed to address specific training needs for data centre and research staff. Staff are also participating in relevant training on research data management and digital preservation including that offered by The National Archives. The BGS also holds the UK Investors in People accreditation that embodies appropriate professional development strategies.

The Head of the NGDC and the Chief Digital Officer are responsible for internal governance of the data centre, and report to the BGS Senior Management Board for the purposes of delivering the agreed services and functions that NERC and BGS expect from the NGDC and the Digital Challenge area.

During 2016, a stakeholder survey was conducted as part of the initial commissioning process for the NERC data centres to evaluate the services that each of the data centres, including the National Geoscience Data Centre must provide to its designated community for the future.

The results of this survey have been used to guide NGDC priorities and the services delivered to users. The outcomes of this process also formed the basis for planning future stakeholder engagement activities that 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). Following the publication of the new BGS Digital Strategy <https://www.bgs.ac.uk/download/bgs-digital-strategy-2020-2025/> and the data centre restructuring exercise, which is currently undergoing an internal consultation, a Data Review Committee has been formed to undertake reappraisal and ingestion assessments in more complex cases where user requirements are not entirely clear.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

6. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository
Accept.

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
Accept

Response:

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.

NGDC staff also proactively engages 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 aligned with current best practices. This includes bodies such as the Research Data Alliance (RDA) where NGDC staff are members, task leaders and co-chairs for a range of relevant interest groups and working groups e.g. Active Data Management Planning IG; Professionalising Data Stewardship IG; the metadata interest and working groups, etc. BGS staff also actively contribute to relevant international initiatives such as CODATA and the Group on Earth Observations (GEO) Data Working Group including co-chairing the newly formed In-situ Data Working Group. https://earthobservations.org/data_wg.php

NGDC staff are also members of relevant professional associations such as the Information and Records Management Society (IRMS).

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:

- Research Data Alliance (RDA) Technical Advisory Board

<https://www.rd-alliance.org/about-rda/our-leadership/rda-technical-advisory-board.html>

- RDA Professionalising Data Stewardship Interest Group

- AGU Data Management Assessment Advisory Board

<https://www.agu.org/Learn-About-AGU/About-AGU/Governance/Committees/Data-Board>

- GEO In-situ subgroup (component of the GEO Data Working Group)

- Information and Records Management Society (IRMS) <https://irms.org.uk/default.aspx>

- Digital Preservation Coalition (DPC) <https://www.dpconline.org/>

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
- DAMA UK <https://www.dama-uk.org/> : a community of data professionals in the UK who champion the value of data management

NGDC staff actively engage in knowledge exchange with a range of organisations, including other repositories, data centres, and similar organisations around the world, e.g. Australian Research Data Commons (ARDC), and the Earth Science Information Partners (ESIP) and 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 feedback received from the NERC survey to evaluate the services each of its data centres delivers to its designated community, the NGDC has undertaken further stakeholder engagement activities. For example, a digital stakeholder survey has been carried out, which provided feedback that will be used to enhance our service and training delivery, and for developing NERC Environmental Data Service (EDS) case studies for reuse of NGDC data. Further stakeholder engagement activities are planned for the future, some of which will be undertaken in collaboration with the other NERC data centres. These activities 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> and other forms of user engagement, for example via social media. In addition, an external advisory board is currently being established that will provide guidance and comment on the activities of both the BGS Digital research area and the NGDC (see also R5 above). The organisation of this panel including potential members has been taken into consideration as part of the planning for the data centre restructure.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase
Accept.

Reviewer 2

Comments:

3 – The repository is in the implementation phase
Accept

Response:

The NGDC has a data ingestion policy in place that sets out the requirements and terms and conditions for depositing data with the data centre. It includes the need to provide complete UK Government GEMINI standard compliant metadata before the data can be ingested.

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/>

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 are 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 unique persistent identifiers are used to track a deposited dataset and ensure internal system integrity, and also feed into NGDC data citation procedures that are documented on the repository website

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-cited-data/?undefined=undefined>.

The depositor of a data set can request a digital object identifier (DOI) which is minted by NGDC using the DataCite service provided by the British Library. A DOI will only be allocated if specific rigorous data management criteria have been met. These requirements are documented on the NGDC and NERC websites:

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-cited-data/> ,

<https://nerc.ukri.org/research/sites/data/doi/>

<https://nerc.ukri.org/research/sites/data/doi/data-citation-guidelines/>

UK Research and Innovation (UKRI) is an organisational member of ORCID that provides unique persistent digital identifiers for researchers. <https://orcid.org/members/001G00000115n0oIAB-uk-research-and-innovation>. UKRI encourages use of this persistent digital identifier across its component research centres and data repositories, including NGDC. ORCIDs are an important element in connecting researchers with their outputs, and also ensure that data centres such as NGDC are able to unambiguously identify data depositors. NGDC will implement mandatory ORCIDs for data depositors that are working on NERC grants, and also provide the facility to include an ORCID identifier for other data depositors in the next year.

The addition of checksums and automatic fixity checks on deposited data are not yet in place, but selected tools are currently being evaluated, and will initially be introduced to the ingestion workflow for new and high priority datasets. 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 and to support audit trails and rollback to a previous state. Following the introduction of fixity checks we will also enhance capture and storage of preservation metadata for data not stored in the database, to allow monitoring and tracking of any intentional and unauthorized changes to data files and replacement of corrupt or altered files where necessary.

We are collaborating with the NERC Environmental Data Service to develop a shared ingestion portal for the transfer of larger NGDC (>1TB) datasets that form part of our catalogue but are stored at the Centre for Environmental Analysis (CEDA) who maintain the JASMIN infrastructure. CEDA generate checksums and conduct fixity checks on these large datasets as standard.

The British Geological Survey has also adopted the International Geo Sample Number (IGSN) system and is working towards becoming the primary UK allocating agent for these unique persistent identifiers for physical samples. <https://www.igsn.org/>. Use of IGSN identifiers creates a permanent link with other associated research assets such as analytical data and the adoption of this system of persistent identifiers for physical material will also provide a link to any associated data held by NGDC where this exists.

NGDC staff are also members of the IGSN steering committee that forms part of the organisational governance structure, which provides a further opportunity to engage with other repositories for the purposes of knowledge exchange.

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:
Accept

8. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accept.

Reviewer 2

Comments:
3 – The repository is in the implementation phase
Accept

Response:

Data deposits are appraised against the Data Collection Policy, the Data Value Checklist (<https://www.bgs.ac.uk/download/ngdc-data-value-checklist/>) and an Ingestion Metadata Policy. Following the Data Ingestion workflow, the data and metadata are checked for completeness. The NGDC maintains a list of acceptable data formats (<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/#guidelines>) which is updated annually based on stakeholder requirements and the US Library of Congress Recommended Formats list <https://www.loc.gov/preservation/resources/rfs/TOC.html>. Any data supplied in other formats requires a dialogue with the donor, and supplementary documentation and metadata about the formats and software may be required. During this process the data can be rejected and sent back to the depositor, or they may be asked to provide further specific information e.g. incomplete metadata, proprietary instruments, software and versions used to generate the deposited data. A standardised workflow is being prepared to ensure consistent metadata capture for other formats to support reusability and preservation in line with the preservation policy <https://www.bgs.ac.uk/download/bgs-digital-preservation-policy/?undefined=undefined>. The NERC environmental data centres maintain a list of other recommended repositories for data that does not fall within their mission/collections profile. The new workplan of the digital data preservation task will assess the entire lifecycle of research data best practice from pre-ingest to preservation and possible disposal of records.

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 the data can be submitted via an online ShareFile platform or the new EDS Large Data Portal which is currently being beta-tested. For those larger data deposits, metadata is still provided via the NGDC portal. Deposits generated using NERC-funded grants are also validated against their active data management plans.

A Deposit Form is required for all submissions, and data will not be accepted unless this is provided. The form 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 in the BGS Data Catalogue that complies with the ISO standard 19115:2003 for geographic information metadata <https://www.iso.org/standard/26020.html>.

Following the publication of the new BGS Digital Strategy <https://www.bgs.ac.uk/download/bgs-digital-strategy-2020-2025/> and data centre restructuring exercise, which is currently undergoing an internal consultation process, a Data Review Committee will be formed to undertake reappraisal and ingestion assessments in more complex cases where individual user requirements are not entirely clear. New policies and procedures are being formulated that may include possible disposal of analogue data after digitisation, where this is not prevented by existing legislation or other obligations as a Place of Deposit, and systematic reappraisal of legacy collections where GEMINI compliant metadata may not be available. As part of the new reappraisal process we are creating a work plan and have identified key criteria for disposal of items from collections.

NGDC is currently working on further developments to enhance the ingestion workflows. It is worth noting that data received via the NGDC Data Deposition Portal has a standardised set of mandatory fields to capture the metadata and data cannot be submitted to us without this information. Consistency of the metadata is checked but not automatically. Due to limited resources and due to the heterogeneity of the geological data (and associated metadata) received, the automation of compliance and format checks is currently a low priority for the NGDC.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

9. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase

Accept.

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Accept

Response:

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).

Following the introduction of the BGS/NGDC Preservation Policy

<https://www.bgs.ac.uk/download/bgs-digital-preservation-policy/>, the internal procedures and the number of copies maintained have been reviewed and updated where necessary. This includes ensuring the original data is stored as it arrived, and that storage and delivery copies that may be edited to improve metadata or migrated to a different delivery format are stored separately from the original data.

Guidance on the transfer of data to NGDC is provided via the BGS website

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/>. Using OAIS terminology, the original data (Submission Information Packages (SIP)) is virus checked before being stored on the local data storage 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 (Dissemination 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 Spectrum Protect (further details of the security procedures implemented within the data Centre are described in Section R16).

Large volume data (>1TB) is ingested to the NERC Environmental Data Service shared data portal hosted by CEDA (Centre for Environmental Data Analysis) who maintain the JASMIN infrastructure. The portal is currently accessed at <https://arrivals.ceda.ac.uk/intro/>. CEDA create checksums for all data ingested through this portal. Work is ongoing to finalise the workflow between the NGDC and the portal, as well as associated user guidance, and to develop a shared NERC EDS branding for the site. All NGDC data deposited via the EDS portal is included in the NGDC data catalogue which points users to the storage location ensuring that all NGDC data is discoverable in one place.

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 Spectrum Protect.

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 BGS Preservation Policy

<https://www.bgs.ac.uk/download/bgs-digital-preservation-policy/?undefined=undefined>, which includes the ISO14721 (OAIS model) for storage as well as the other preservation functions. Checksum and fixity check procedures for new data ingested at the NGDC will also be initiated in the near future, and work is ongoing to design the workflow and to assess the priority data for trialling this procedure.

NGDC/BGS use Linear Tape-Open (LTO) to store backups and archives. During 2019/2020, SNS have migrated our backup and archive data to LT07 tapes in a new Hewlett Packard Enterprise library (model MSL6480). The LT07 tapes can store up to 15TB using encryption, have an archival life of 30 years and with an expected 20,000 load/unload cycles.

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 faulty tape is securely destroyed. All old or damaged tapes are securely destroyed on site by a third party using a crushing machine.

Daily logs are produced by the IBM Spectrum Protect 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

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

10. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accept.

Reviewer 2

Comments:
3 – The repository is in the implementation phase
Accept

Response:

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

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/> In an article based on a MSc dissertation in 2016 (<http://nora.nerc.ac.uk/id/eprint/517250/>) one of the key findings was that the validity of geoscience data held at the NGDC is much longer than 10 years, and where the work was funded by the UK Government the data may be held in perpetuity under the Public Records Act. These two points take precedence over the UKRI 10-year retention.

The NGDC has introduced a dedicated programme to manage long-term digital preservation and to assume a stewardship role that ensures the survival of digital materials from one generation of digital systems to the next. NGDC published its first digital preservation policy in 2017, and this was updated in 2020

<https://www.bgs.ac.uk/download/bgs-digital-preservation-policy/>. It now also maintains an accompanying digital preservation strategy and work plan. In addition, NGDC has recently completed the Digital Preservation Coalition Rapid Assessment (RAM) exercise (<https://www.dpconline.org/our-work/dpc-ram>) to benchmark its digital preservation maturity and capability. NGDC has also run an internal digital research data survey to investigate researchers' data management practices. The resulting internal report (IR/20/003, available on request) explores the impact of existing practices on the

long-term resilience of data assets stored within the NGDC and makes recommendations on data centre service development. The follow-up work will include building a digital information asset register dashboard and undertaking a risk assessment aligned with the new BGS Digital Strategy. It will take into account the level of preservation for different data collections and digital objects based on the requirements of the designated community, 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/>) for the BGS Data Catalogue, 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. This will initially be captured for a subsection of key ingested datasets and also for large volume datasets transferred to the NERC Environmental Data Service Archive hosted by CEDA (Centre for Environmental Data Analysis).

The NGDC 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 is archived including the corporate work tasks and integrity checks used to migrate the entire database to the new version. Both the original formats/bit streams and any converted data are also 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 the 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, good data deposit guidelines, 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. The associated 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 fully 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. NGDC provides PhD students with regular domain-focused

training to ensure they gain the relevant professional research data management skills, and eventually deposit robust datasets with the NGDC as part of their NERC-funded activities. This training course and its anticipated long-term impact on NGDC data assets is described in an iPRES 2019 conference paper (<https://osf.io/cr2ah/>).

BGS has been a partner in several relevant EU-funded projects including the previous SciDIP-ES and the ongoing ENVRI-FAIR that are addressing different aspects of data sharing including domain level adoption and implementation of the FAIR data principles. NGDC staff are also active participants in relevant European and international initiatives such the Research Data Alliance (RDA) and the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS) with the aim of expanding and developing their knowledge and expertise in the field of data stewardship. The benefits derived from involvement in these activities includes awareness of innovative tools and methodologies that can be implemented within the NGDC to improve the effectiveness of its data management processes.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

11. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase
Accept.

Reviewer 2

Comments:
3 – The repository is in the implementation phase
Accept

Response:

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. The Portal provides guidelines for depositors e.g. on good data deposits and acceptable file formats <https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/?undefined=undefined>. Deposits relating to particular areas such as NERC Grants, and internal BGS data 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. Where possible, data centre staff liaise with data providers to rectify issues identified at the data ingestion stage, e.g. to request additional documentation or metadata.

Discovery metadata for the datasets held by the NGDC is made available via the BGS Data Catalogue: <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 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 Service (EDS), one of which is the NGDC, to check the scientific quality of all the diverse geoscience data it receives. However, the EDS has processes in place to ensure that the data supplied has appropriate metadata to describe the data asset, and data files have headers, units of measurement, consistent population and undergone other quality checks. Data generated by NERC funded grants will be reviewed and validated against NGDC data management plans where these are mandated under the NERC Data Policy.

BGS Enquiries Service provides a direct feedback mechanism for data users to comment on data quality, availability, systems, data formats, missing or incomplete data, and other issues either by emailing enquiries@bgs.ac.uk or phoning the Enquiries team, who then feed the information back to the NGDC team using an Issues Log that is regularly reviewed.

The NGDC is embedded within the British Geological Survey, and can draw on the extensive expertise and knowledge of its 650+ staff, which include recognised UK and international experts in the geosciences, as well as data and information management and IT/IS infrastructure and development professionals. This direct access to the large and diverse pool of expertise within the hosting research centre is a significant added benefit and strength for NGDC, which would not be possible if it were a standalone facility.

As the funding agency, UKRI-NERC commissions projects from a diverse range of academic and research organisations. These commissioned projects are expected to generate high-quality research assets including data, papers, models,

code, software, etc., which adhere to the relevant guidance from both ourselves (as the data centre) and their home institutions. Any resulting data submitted to the NGDC is checked to ensure it complies with appropriate formats and includes the necessary headers, units, and classifications before it can be accessioned and processed. These checks also ensure the validity of the associated metadata as well as the robustness and usability of the individual data files.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

12. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository
Accept.

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository
Accept

Response:

Repository comment on 1st February 2022: "The workflows for supporting accession and ingestion are encapsulated within our NGDC online data portal and supporting registration and delivery systems which provide links to documents, best practice, and guidance all available externally.

We are happy with a rating of '3' if both reviewers consider that the portal and external guidance/policies/checklists we provide are not sufficient in meeting these criteria at the higher rating '4', this is presently not the case (one '3' and one '4' from the reviewers). We can work on additional workflow documents for future years to explain the internal workflow that is

encapsulated in the portals and systems used by the Data Centre staff.

At the October 2021 review both reviewers accepted a rating of '4' and this changed to the January 2022 review - criteria are being re-evaluated differently each time the case is considered. We didn't expect to need to re-do/supply more information for this criteria when it was previously agreed and there were no reviewers comment to address."

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

All NGDC data acquisitions must adhere to the relevant NGDC data collection policy

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/> and data value checklist

<https://www.bgs.ac.uk/download/ngdc-data-value-checklist/>. The ingestion process begins by ensuring the data submission complies with both of these requirements. The process is implemented using a fully documented 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

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/?undefined=undefined#confidentiality> .

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 in consultation with the depositor. Data depositors are guided through the submission process via the NGDC webpages

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/#guidelines> . 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.

Data ingested through the online digital NGDC Data Deposit Portal into the corporate Detailed Accessions Database (<https://webapps.bgs.ac.uk/services/ngdc/accessions/index.html>) receives an internal unique identifier. Data depositors are also offered the possibility of obtaining a DOI for their datasets on request

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-cited-data/>. 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 UK DPA 2018, FOIA 2000 <https://nerc.ukri.org/about/policy/foi/> , EIR 2004 and PRA 1958/1967 legislation. These are included in the UKRI Records Management Policy <https://nerc.ukri.org/about/policy/foi/records-management-policy/> and the NERC Data Policy <https://nerc.ukri.org/research/sites/data/policy/>

The NGDC has a digital preservation policy that is published on its web pages <https://www.bgs.ac.uk/download/bgs-digital-preservation-policy/> . An accompanying strategy and a work plan (as described under section R10) have been produced that includes appropriate annually reviewed 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 datasets, ensures transparency throughout the data selection and archiving process.

The NGDC is part of the NERC Environmental Data Service (EDS) <https://nerc.ukri.org/research/sites/environmental-data-service-eds/> that holds environmental data on behalf of NERC and makes them openly available and accessible for users. The EDS develops and delivers integrated and shared services such as the NERC Data Catalogue and the EDS Large Data Donation Store. There are plans to collaborate more widely with the other NERC data centres on digital preservation activities, and to share the expertise the NGDC has already gained in this field.

The NGDC provides online/email and help desk support for data depositors. A variety of documents are also available online to help and guide data submission, which include guidance on a good deposit, data value checklists and list of preferred data formats. These resources are in addition to the direct interaction we undertake for larger scientific programmes where NGDC staff hold sessions and meetings to explain data and information management issues, challenges and explore best practice, which help a project or programme to create and supply appropriate scientific research data. Our online data deposit portal encapsulates workflows to guide depositors in supplying the necessary metadata, terms and conditions of use, and data files to ensure a robust and comprehensive data deposit. The data centre has internal systems to process the accession and ingestion of the data, create formal metadata records, move files to appropriate storage locations and enable subsequent online access.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

13. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:

3 – The repository is in the implementation phase
Accept.

Reviewer 2

Comments:

3 – The repository is in the implementation phase
Accept

Response:

NGDC provides direct access to a range of key datasets through its OpenGeoscience service

<https://www.bgs.ac.uk/geological-data/opengeoscience/> 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

<https://metadata.bgs.ac.uk/geonetwork/srv/eng/catalog.search#/home> . 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 has a documented data citation process for selected “approved” datasets, which are available on its website <https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-cited-data/?undefined=undefined>.

Information is provided on the rationale that has been adopted for selecting these ‘approved’ datasets, including the associated data citation process.

The NGDC also makes available a number of “approved” 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 available for re-use following citation.

The NGDC metadata is compliant with Gemini V2.3, a UK national standard and a derivative of ISO 19115. It is mandated by our funder, UKRI-NERC and the UK Government. This metadata is held within an appropriate relational database and externally accessible on a C-SW (GeoNetwork) service allowing harvesting by metadata aggregators including at this time, the NERC Data Catalogue (our parent organisation), Data.Gov.UK (UK government metadata portal), JRC Portal (European data portal) as well as the catalogues for several research infrastructures.

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:

Accept

14. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept.

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Response:

Discovery metadata complying with ISO standards 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 allows datasets to be exposed through appropriate external gateways (e.g. data.gov.uk, and the NERC data catalogue).

The NGDC publishes a list of preferred file formats for deposit at

<https://www.bgs.ac.uk/geological-data/national-geoscience-data-centre/ngdc-depositing-data/> in an effort to encourage deposit of data in formats that support open access, are at less risk from technology/software obsolescence, and provide efficient migration paths to newer file formats when necessary to ensure digital continuity. These 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., generic formats such as CSV, TXT, and in some instances other database formats for raw data. Spatial

data is typically submitted in ESRI ArcGIS formats. The list of preferred data formats is reviewed annually in conjunction with updates to the Libraries of Congress Recommended Formats Statement <https://www.loc.gov/preservation/resources/rfs/> and taking into account any changes in the requirements from the designated community. Data deposited using preferred formats can be ingested routinely; for other formats, especially proprietary ones, additional metadata may be required. A standard ingestion process is being developed to ensure consistency in harvesting this information.

A modular digital preservation programme (both policy and strategy) has been introduced to ensure the continued usability of the data, which is a key element of the NGDC data management procedures. The preservation strategy includes 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. For example, data provided in older spreadsheet formats (e.g. earlier versions of Microsoft Excel, or obsolete Lotus formats) which has been converted into more resilient CSV, TXT, or ASCII formats.

In order to ensure continued understandability of the data, appropriate contextual metadata and other associated documentation (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 mandatory fields.

The NGDC is also frequently required to store a range of datasets generated by geoscience research and other environmental modeling. In order to encourage the re-use of this data, a rich level of contextual metadata has been 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. Model code may be offered to the NGDC who stores it on NGDC GitHub <https://github.com/britishgeologicalsurvey-ngdc>. A shared NERC EDS model portal is in development as part of data centre integration activities which is currently undergoing beta testing. The new portal will be more user friendly and allow researchers to input the model code to the portal directly.

The NGDC recognizes the importance of providing as much contextual information as possible for its data holdings. 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.

As one of the key partners in the UK Geospatial Commission, NGDC has initiated a user engagement activity that will provide knowledge and expertise beyond the current designated community. It will also inform future strategies for stakeholder engagement, systems and work plans at the NGDC to encourage wider adoption of best practices and good data management.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

TECHNOLOGY

15. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository
Accept.

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository
Accept

Response:

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. It is reviewed by them to ensure that Data Centre requirements are covered and is due for upgrade to a larger capacity and faster system in 2021. Data that does not require on or near-line storage is archived to Enterprise Tape Libraries, implemented using IBM Spectrum Protect, 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 Spectrum Protect 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 upgraded dual-redundant 10Gbps links providing 24 hour a day access to NGDC data and maximizing service availability.

Ingestion

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).

The NGDC shares its infrastructure with the British Geological Survey, a long-standing geological survey, who have been

building digital systems since the 1970's. BGS is a research centre that is part of the UK Research and Innovation (UKRI). The complex and well established integrated relational database (Oracle) is core of the NGDC/BGS shared architecture and includes comprehensive standards compliant metadata and has been built and enhanced over many decades. Extensive use of unique persistent identifiers, ISO dictionaries and domain specific scientific classifications and vocabularies ensure the data maintains data integrity and linked files are easily found from comprehensive metadata. The architecture supports the research centre, the data centre, other science facilities and collaborations with the geoscientific community as well as research infrastructures (such as EPOS).

The data infrastructure is based around two primary sites, Nottingham (Keyworth campus) and Edinburgh (co-located with Heriot-Watt University). Both have multiple high-speed WAN connections with compute and storage capacity. Replication between sites is undertaken in order that if one site was unfortunately taken down the workforce, primary systems and data services could be run from the secondary site in a short time frame. This is managed by F5 load balancing solution which enables user/traffic to be directed to an appropriate site infrastructure. Not all NGDC data can be replicated over the multiple sites, that is a future ambition but we have tape backups both on-site and off-site copies and access to capacity in the NERC large data storage service, based within JASMIN/CEDA at RAL.

The BGS/NGDC infrastructure and architecture is enhanced in accordance with funding received to meet requirements gathered from its user base (the NGDC being a primary user) to deliver the best and most robust solution achievable within the funding available. This has recently included upgrades to our data warehouse, additional API interfaces to key datasets, entity-relationship modelling software, WAN bandwidth increases to both primary and secondary access pipelines and a new higher capacity SAN system.

This solution enables our scientific domain/community to better engage, access and query the geoscientific data held and encourages re-utilisation. It does not rely upon an old generic standardised repository model which would only offer a simple query method to objects held.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

16. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository
Accept.

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository
Accept

Response:

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. Microsoft Active Threat Protection is running on the SAN nodes as well as all End Points to protect against any malicious activity. 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 Spectrum Protect. 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. 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 recently received the Cyber Essentials accreditation (<https://www.ncsc.gov.uk/cyberessentials/search>), and attained the ISO:9001:2008 Quality Management Standard for its quality management systems.

Further work is being carried out to enhance information security training that will be rolled out in the coming months which will enhance our capabilities in this area.

A BGS Risk Register has been drafted pending completion of a restructuring of the Digital Area of the organisation and the confirmation of revised key roles. Once this has been completed, the creation of a final version of the Risk Register will

completed, identifying priority datasets and also any areas of data risk for consideration and mitigation.

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

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These Requirements are not seen as final, and we value your input to improve the CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response:

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. <https://www.bgs.ac.uk/about-bgs/our-work/accreditations/>

It should be noted that BGS has recently undergone a restructuring process, which has repositioned the NGDC at the core of a number of organisational functions. However, as part of this process there has been an associated redesign of the BGS and the NGDC websites that has impacted the online availability of NGDC documentation. For this reason, while all links in this document have been verified at the time of submission, some links may be broken due to ongoing changes to the website. If any documentation is unavailable via links in this document it can be obtained directly from the data centre by emailing enquiries@bgs.ac.uk

The difference in the reviewers' comments and also their level of scoring is significant and demonstrates a need for standardisation and better understanding across the pool of reviewers. We have no issue with any reviewer challenging our re-certification or comments but do have concerns when one reviewer says '3' and says its possibly not scorable due to lack of evidence and the other says approved '4'.

There was always an understanding that not all documentation can or would be made public for all parts/workflows of any data centre or repository, some could be provided offline for reviews to consider, others may not be released, this was understood by CTS and acknowledged. It does not seem to be basis of some comments received. Has this policy changed and not been disseminated to the community?

The value and USP of CoreTrustSeal was its pragmatic place in the certification marketplace. It needs to hold the centre ground, encouraging continuous development and improvement while also maintaining its applicability to the wider range of repositories and data centres. Taking care not to be as onerous as ISO (and therefore if little interest or affordability to the many) but equally also not a subscription and simple tick box exercise. Care is needed to ensure that the criteria and especially the assessment of them doesn't lift CTS too far up towards ISO or remove the key benefit of CTS which is its applicability to wide range of repositories and data centres.

Reviewer Entry

Reviewer 1

Comments:

Note from CoreTrustSeal Board: It is evident from your application that you are in transition. Your application is accepted this time but in your next round the board would expect to see more requirements at level 4.

Reviewer 2

Comments:

Note from CoreTrustSeal Board: It is evident from your application that you are in transition. Your application is accepted this time but in your next round the board would expect to see more requirements at level 4.