Assessment Information

[CoreTrustSeal Requirements 2020–2022](#)

Repository: UK Polar Data Centre
Website: https://www.bas.ac.uk/data/uk-pdc/
Certification Date: 18 March 2021

This repository is owned by: **British Antarctic Survey**

UK Polar Data Centre

Notes Before Completing the Application

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

True

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:

Domain or subject-based repository, Institutional repository, National repository system; including governmental

Brief Description of Repository

The UK Polar Data Centre (PDC) is the UK national data centre for polar and cryospheric sciences. It is hosted at the British Antarctic Survey (BAS) and funded via its parent organisation, the Natural Environment Research Council (NERC), which is part of UK Research and Innovation (UKRI, www.ukri.org).

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

The PDC holds polar and cryospheric data assets primarily from NERC funded polar research grants and programmes as well as those created by the Survey's own scientific programmes. It provides long-term curation and public dissemination of these assets.

Brief Description of the Repository's Designated Community.

The designated community for the PDC consists of a wide range of users of polar and cryospheric data. Users are international and typically in academia, government departments and agencies (e.g. UK Foreign and Commonwealth Office, FCO, and Department for Environment, Food and Rural Affairs, DEFRA), environmental non-governmental organisations (e.g. Birdlife International), Governments of UK overseas territories (e.g. Government of South Georgia and the South Sandwich Islands) and other international bodies such as the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), the Scientific Committee for Antarctic Research (SCAR) and the Southern Ocean Observing System (SOOS).

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Comments

The PDC aims to preserve and disseminate its data holdings for the long-term. The data held by the PDC provides an evidence base for scientific programmes, underpins scientific research and is a source of critically important scientific data for inclusion in future scientific projects, applications, proposals and decision support systems. The PDC facilitates scientific peer-review processes, open access and re-use of internationally important datasets and information products.

A UK national Antarctic data centre, called the Antarctic Environment Data Centre was set up in 1992 to allow the UK to meet the Antarctic Treaty recommendation that every consultative party should establish a National Antarctic Data Centre. Following a review of its services in 2008, its remit was expanded to cover all polar regions, as well as cryospheric data, and its name changed to the UK Polar Data Centre. The data held by the data centre have been derived from over 60 years of polar research carried out at the British Antarctic Survey, and other UK institutions.

Information on the PDC can be found on its website (https://www.bas.ac.uk/data/uk-pdc/). This includes information on how to find, deposit and cite data, the data policies we apply and the support that we can provide.

The PDC undertakes a variety of levels of data curation. For long-term datasets collected as part of BAS long-term monitoring programmes, the PDC is involved in the whole process from the point of data collection (e.g. from instruments on Antarctic bases or ships) right up to the long-term archiving of the data. Data managers will monitor the data recorded by instrumentation, ensure it is transferred securely to our data storage systems in Cambridge and undertake quality control and data processing tasks (see https://www.bas.ac.uk/data/uk-pdc/operational-data-management/). For other

datasets, the PDC receives data at the end of completed scientific projects and undertakes a mixture of basic and enhanced curation tasks. Depositors are supported to produce quality metadata to describe the data they are depositing (see https://www.bas.ac.uk/data/uk-pdc/metadata-guidance/) and PDC data managers check the metadata provided and discuss any changes required with the depositor. The PDC has a list of preferred data formats (https://www.bas.ac.uk/wp-content/uploads/2020/06/PDC_Data_Format_Guidance.pdf), but will support depositors in converting their data from a proprietary format to an open one if required.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

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

The PDC is hosted by British Antarctic Survey (BAS) and is dependent on the BAS IT infrastructure. It is one of five Environmental Data Centres funded by the Natural Environment Research Council (NERC) as a single Environmental Data Service. The baseline functions that all NERC data centres are expected to deliver have been defined and a document outlining these has been provided for the reviewers as it is not publicly available. NERC is one of the seven UK research councils that form UK Research and Innovation (UKRI).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

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

Not Applicable

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Reviewer Entry

Reviewer 1

Comments:
Accept

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

Reviewer 2

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

Response:

The UK Polar Data Centre (PDC) is the focal point for Arctic, Antarctic and cryospheric environmental data management in the UK. Part of the Natural Environmental Research Council's (NERC) network of environmental data centres (https://nerc.ukri.org/research/sites/data/) and based at the British Antarctic Survey (BAS), we coordinate the management of polar and cryospheric data from UK-funded research and support researchers in complying with national and international data legislation and policy.

The PDC enables the UK to meet the Antarctic Treaty data obligation, Article III (1) (c) that "scientific observations and results from Antarctica shall be exchanged and made freely available" and the objective of the International Arctic Science Committee (IASC) to "seek to ensure that scientific data and information from the Arctic are safeguarded, freely exchangeable and accessible". The PDC's scope and mission is clearly stated on our website at https://www.bas.ac.uk/team/business-teams/information-services/uk-polar-data-centre/.

We recognise the value of scientific data and the importance attached to the long-term professional management and preservation of data assets. Data are vitally important both as an evidence base for existing scientific projects and for future re-use. Our work enables scientists to comply with the NERC Data Policy (https://nerc.ukri.org/research/sites/data/policy/data-policy/) and the common principles on data policy of our funding body UK Research and Innovation (https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/).

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:

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

Reviewer 2

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

Response:

The PDC makes most data available under the terms of the UK Open Government Licence (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/). Other open licences may be used for international data collaborations that are hosted by the PDC such as Creative Commons Attribution 4.0 International (https://creativecommons.org/licenses/by/4.0/).

Depositors are asked to accept the use of the UK Open Government Licence in the Data Transfer Agreement (https://www.bas.ac.uk/wp-content/uploads/2019/06/Data-Transfer-Agreement.docx), unless agreed otherwise. Depositors must ensure compliance with their funders requirements, which may dictate the licence used for a particular dataset. Most data received by the PDC originates from NERC funding and the depositor must therefore comply with the NERC Data Policy (https://nerc.ukri.org/research/sites/environmental-data-service-eds/policy/datapolicy-guidance/) and use the UK Open Government Licence where possible.

The PDC does not normally accept data where there is a disclosure risk and asks that any data containing sensitive information be anonymised or be of a sufficiently broad scale that sensitive location information is not made publicly available. Should the PDC be required to publish sensitive information, the licence used will reflect the sensitive nature of the dataset and restrict access appropriately.

Licensing information and any required citations/acknowledgements are included in the metadata record describing each dataset in the use constraints section (e.g. https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/01226).

We currently have no means of identifying non-compliance with any licences unless it is brought to our attention by others. The most likely form of non-compliance with the open nature of the licences used would be someone using the data and not giving appropriate citation. The scientific review process ought to identify anyone including such data in publications without attributing the appropriate source. If, however, we become aware of breaches of licence or access conditions then we contact the offending user to raise the issue. If this does not lead to a satisfactory resolution then the issue would be raised with the data owner for them to pursue further.

It should be noted that the Freedom of Information Act 2000 (FOIA), Environmental Information Regulations 2004 (EIR), Public Records Act 1958/67 (PRA) and the General Data Protection Regulations 2016 (GDPR) may override any licensing agreements.

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

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The PDC is implemented as an integral part of the BAS infrastructure where it is part of the Innovation and Impact Directorate. The BAS Director of Innovation and Impact is the business owner of the PDC digital preservation function and line manages the Head of the PDC. This director sits on all senior BAS committees (including the BAS Executive Team, BAS Management Team and the Science Strategy Team) allowing issues related to data management to be raised at senior levels. The PDC reports monthly to the BAS management team and provides presentations to the senior BAS committees when relevant. A risk register for the directorate is maintained and updated regularly which highlights any risks to the ongoing access and preservation of data. Actions are then taken to mitigate any risks identified.

Being part of the BAS infrastructure enables the PDC to utilise the data storage, systems and networks provided by the BAS IT department and includes high-speed access to the UK Joint Academic Network (JANET). BAS operates under the UKRI business continuity management policy (this policy is attached for reviewers) and business continuity plans are constantly under review with annual desktop exercises to test these plans being designed for the future. As PDC data assets are held within, and are accessible from, the BAS infrastructure, their future continuity is predominantly determined by that of the organisation.

However, the PDC is also one of the five NERC data centres which have been commissioned by NERC as a single environmental data service (currently for a five year period until 2023 – see section 5). The data centres are working together to increase integration and use of common processes and software tools. As part of this integration there will be a move to use the JASMIN super-computing centre at Harwell (funded by UKRI) for the long-term archiving of large datasets. This integration provides increased confidence in longevity of data held by the PDC as it is no longer solely reliant on the infrastructure at BAS. In the unlikely scenario of BAS no longer wishing to host the data centre then the data could potentially be transferred to one or more of the other NERC data centres or hosted on JASMIN.

A digital data preservation policy for the PDC is available publicly through the PDC website (https://www.bas.ac.uk/wp-content/uploads/2020/05/PDC_digital_data_preservation_policy_v1.0.pdf). This policy has been developed using best practise for the UK community and describes the reference model that PDC has adopted and includes sections on preservation planning, IT infrastructure and security, funding and resource planning.

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

Response:

The PDC complies with the NERC Information Security Policy and NERC Data Protection Policy to safeguard and protect the information and assets in its care. For security reasons these are internal documents that are only available to staff (a copy of the NERC Information Security Policy has been provided for reviewers). All staff working in the PDC are required to undergo regular training on Data Protection and Cybersecurity. This training ensures that all PDC staff are aware of issues related to personal/sensitive data.

The PDC also complies with UKRI and NERC policies (https://nerc.ukri.org/about/policy/policies/). These provide 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 have the authority to take any necessary action.

The majority of our users who deposit data are funded by NERC and also bound by relevant NERC policies and guidelines. NERC publishes a Research Grants and Fellowship Handbook (http://www.nerc.ac.uk/funding/application/howtoapply/forms/grantshandbook/) that includes guidance on research ethics. Researchers are required to comply with the UKRI 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/). These policies and guidelines apply equally to researchers, support staff, research administrators, Research Council staff and all individuals contributing to the Research Councils' peer review process.

The PDC complies with the Freedom of Information Act 2000 (FOIA), Environmental Information Regulations 2004 (EIR), the Public Records Acts 1958/1967 (PRA) and the General Data Protection Regulations 2016 (GDPR). These legislative requirements are included in the NERC Records Management Policy (https://nerc.ukri.org/about/policy/foi/records-management-policy/) and NERC Data Policy (https://nerc.ukri.org/research/sites/data/policy/data-policy/)

In cases of non-compliance with these conditions, the UKRI can invoke its disciplinary policy to ensure that the highest standards of behaviour and conduct in research are met (https://www.ukri.org/files/termsconditions/rcukukriterms/disciplinary-pdf/).

The PDC has an ingestion process in place (https://www.bas.ac.uk/data/uk-pdc/data-deposit/) that ensures any disclosure risks are considered at an early stage in the process. Depositors are required to sign a Data Transfer Agreement (https://www.bas.ac.uk/wp-content/uploads/2019/06/Data-Transfer-Agreement.docx). This agreement asks for assurance that the data have been collected within ethical norms and are legally appropriate i.e. guaranteeing that nothing in the resource or supporting information contravenes the Data Protection Act 1998 or any other EU/UK law.

Once data are received from the depositor (and prior to them being made publicly available) a number of acceptance tests are carried out by data centre staff, using the Data Quality Checklist (https://www.bas.ac.uk/wp-content/uploads/2020/04/PDC_Data_Quality_Checklist.pdf). These include checks that the data and supporting documentation do not contain any personal information or otherwise contravene GDPR. A record of these checks having been carried out is kept by the data centre in perpetuity.

The PDC does not normally accept data where there is a disclosure risk and asks that any data containing sensitive information be anonymised or be of a sufficiently broad scale that sensitive location information is not made publicly

available. The PDC does not currently hold any data related to the local and traditional knowledge of indigenous people in the Arctic. It follows the International Arctic Science Council's statement on Principles and Practices for Arctic Data management (https://iasc.info/images/data/IASC_data_statement.pdf) which stipulates that where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised.

In the unlikely event that data are made available that contain sensitive information, the PDC will take appropriate action to withdraw these data. The event will also be recorded as a non-conformance on our non-conformance log where we will identify the root cause and put corrective action in place to stop it occurring in the future.

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:

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

Reviewer 2

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

Response:

The PDC is funded through the NERC which is part of UKRI. Part of the funding comes from the NERC National Capability Data Centre Commissioning project, part from the top-slicing of relevant research grants to fund data management and part from BAS core funds. The PDC relies on the underlying BAS IT infrastructure (see section 3). The funding for the

ongoing management and maintenance of the IT systems is provided by BAS and additional bids are made for capital via the BAS budget bidding process and additionally to NERC/UKRI for capital funding to upgrade systems and storage availability. Recently a successful bid has resulted in the procurement of a new central storage solution.

NERC has commissioned its five data centres as the NERC Environmental Data Service for an initial period of five years (until 2023). As part of this process, the services and functions that must be delivered by each data centre, including the PDC, have been defined. The number and quality of these are reported on to NERC annually. Towards the end of the commissioning period, a review of the service will be undertaken prior to its commission for the next five years.

BAS delivers and enables world-leading interdisciplinary research in the polar regions and has been responsible for much of the UK's scientific research in Antarctica for the last 60 years. BAS provides the funds for operational data management so that PDC staff can support scientists in ensuring that data collected in remote polar regions, from research stations, field sites and research vessels are securely and safely returned to backed up file servers at BAS and properly managed for the long term.

The PDC is managed by the Head of the PDC who reports to the BAS Director of Innovation and Impact. There is currently a full time equivalent of 9.25 staff, 1.25 of which are fixed-term project funded contracts. PDC staff are experienced data specialists with a range of skills, including digital preservation, scientific data accession, active data management planning, information architecture, and software development. In addition, we ensure that our data managers have in-depth knowledge of the science they support allowing them to, thoroughly and critically, assess datasets and their associated metadata. Staff have access to the learning and development programme provided to all BAS employees and have time within their forward job plans for professional and personal development. There are also opportunities, and funds available, for staff to attend relevant conferences and workshops to ensure their knowledge stays up to date.

The PDC is involved with relevant national and international committees and associations including the Standing Committee for Antarctic Data Management (https://www.scar.org/data-products/scadm/overview/), the Arctic Data Committee (https://arcticdc.org/), Southern Ocean Observing System data management sub-committee (http://soos.aq/index.php/data/data-committee), the NERC Information Strategy Group and Data Operations Group (see question six).

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

Response:

There are mechanisms in place for the PDC to secure ongoing expert guidance and feedback both from the NERC data community and the international polar data community.

PDC staff regularly interact with staff from other NERC data centres (National Geoscience Data Centre at the British Geological Survey, Environmental Information Data Centre at the UK Centre for Ecology and Hydrology, British Oceanographic Data Centre at the National Oceanographic Centre and the Centre for Environmental Data Analysis at the Science and Technology Facilities Council) both on an informal basis and formally through the NERC Information Strategy Group (ISG) which meets 4 times a year and the Data Operations Group (DOG) which meets 6 times a year. Two of these data centres (National Geoscience Data Centre and Environmental Information Data Centre) already have Core Trust Seal certification.

NERC's Science Information Strategy is led by the ISG which comprises Directors of Institutes, Heads of NERC Data Centres and external advisors.
The Head of the PDC is a member of ISG whose remit is:
1. To develop and maintain a NERC Science Information Strategy and action plans, incorporating information management, acquisition, delivery, systems and technology in support of NERC's objectives;
2. To provide advice and guidance on scientific information issues relevant to the NERC Director of Science; and
3. To advise NERC on information policies, standards and practices needed to deliver the strategies efficiently and effectively.

Strategic direction of NERC environmental data management is provided by the ISG and implemented across the NERC Data Centres through the DOG. Two members of the PDC sit on the DOG.

Internationally, the PDC is part of the Scientific Committee for Antarctic Research (SCAR) Standing Committee on Antarctic Data Management (SCADM), the Antarctic Data Committee (ADC) and Southern Ocean Observing System Data Sub-Committee. These committees all meet annually and in between have regular teleconferences to enable sharing of best practice, setting priorities and development of policies and procedures. Staff also participate in relevant international meetings such as those held by the Research Data Alliance (e.g. I-ADOPT working group, Persistent Identifiers of Instrumentation Working Group and RDA UK group) and World Data Forum (a member of the PDC has just finished her term as co-chair of the WDS Early Career Researchers and Scientists Network) and data relevant sessions of bodies such as the European Geophysics Union.

The PDC is an integral part of BAS meaning there is a very close relationship with many of the scientists collecting UK polar data. Each BAS science programme has a named data manager who has a background in the relevant domain as well as data management expertise (https://www.bas.ac.uk/data/uk-pdc/contacts/). The wider UK polar science community is reached by attendance and presentations at biennial Arctic and Antarctic conferences. These close working relationships allow the PDC to shape its services to meet its designated community. Feedback can also be sent via our web pages (https://www.bas.ac.uk/data/uk-pdc/contacts/) or via our feedback form (https://www.bas.ac.uk/pdcfeedback), which allows anonymous comments from any user.

The NERC DOG has recently established a communications subgroup which is working to develop a stakeholder engagement strategy. A member of the PDC is part of this subgroup. This strategy will target users of the NERC Environmental Data Service to better understand what the science communities are likely to need/want from the service in the future. This will help us to shape future PDC developments/services to ensure we continue to meet our users' needs.

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

Response:

The PDC clearly informs users on what is required in order to deposit data (https://www.bas.ac.uk/data/uk-pdc/data-deposit) and on data policy (https://www.bas.ac.uk/data/uk-pdc/data-policy/). It maintains a catalogue of metadata records, the Discovery Metadata System (DMS, https://www.bas.ac.uk/project/dms/) that is compliant with the ISO 19139 XML implementation schema for the ISO 19115 Geographic Information – Metadata standard. All changes to this catalogue are logged so there is a complete history of edits made to each record.

The metadata required for depositing research data in the PDC are described on our website (https://www.bas.ac.uk/data/uk-pdc/metadata-guidance/) and a template is provided to help users ensure they provide all required information (https://www.bas.ac.uk/wp-content/uploads/2018/12/PDC_Metadata_Guidance_version7.docx). Data are only published once the requested metadata are provided. In cases of insufficient metadata, the PDC data coordinator or relevant data manager will liaise with the depositor to obtain the relevant metadata for compliance with the metadata standards and to enable reuse of the data without recourse to the original author. All depositors have to sign a Data Transfer Agreement (https://www.bas.ac.uk/wp-content/uploads/2019/06/Data-Transfer-Agreement.docx) which clearly states to depositors what the PDC can do to the data to maintain its readability and integrity over time. The identity of depositors is verified by matching contact details provided by the depositor (email address, institutional contact), publication references and the depositor's other online presence, such as entries on Research Gate or LinkedIn. To provide others with confidence in the identity of data generators and collaborators, the DOI XML created by the PDC contains the relevant ORCID for these people whenever they are available. Those without ORCIDs are encouraged to register for one.

The PDC enables data transfer via the secure password-protected BAS FTP service. A project is underway to allow checksums to be used when data are deposited with us (to ensure no corruption during transfer) and subsequently to regularly check that data files in our repository have not changed (see section nine). A test system has been developed which runs a script daily to check the published data area for any changes to which files are present and their checksums. Any changes are logged and a warning message is sent to enable further investigation. This system will be fully implemented within the first half of 2021. Deposited datasets are kept in a separate storage area whilst the accession is in progress. A record of changes to the data from the original deposit to the final published version is kept in several systems, depending on data type (JIRA, Gitlab or email thread archive). Following approval of the final version from the

data depositor, the final dataset is moved to a data area specifically for published data. Once a dataset has been published then no further changes are allowed (https://www.bas.ac.uk/data/uk-pdc/data-citation-and-publishing/obtaining-a-data-doi/). Should a dataset be superseded then the removal of the original dataset from the publicly accessible collection is assessed on a case by case basis. If it is decided to remove immediate access to a dataset (for example, if it contains errors that have been corrected in a subsequent version), the metadata will remain available and the data will be available on request. It is made clear in the metadata that the data has been superseded and a link is provided to the updated version.

A new workflow is being developed for the PDC data accessions which will allow us to keep consistent audit trails of any changes made to metadata or data within a single system rather than using a number of different systems as described above. Currently we are investigating solely using either JIRA or Gitlab for this purpose, both of which are used within the team currently as described above. The review of which system will work best for all data types is ongoing, after which it will be relatively quick to implement given the team are familiar with both systems. The new workflow will be put in place during 2021 and will: log all communication between the PDC and the depositor(s); monitor the ingestion of data; manage any data embargoes; and track updates made to the metadata and data.

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:

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

Reviewer 2

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

Response:

All data deposit requests to the PDC are initially assessed using the NERC Data Value Checklist (https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/) and an internal Data Acceptance Procedure. The former is used to determine the long-term value of the data, along with the integrity and originality of the data. The latter determines whether a particular data resource is within the PDC's remit, which includes data from polar and cryospheric regions as well as other areas within the British Antarctic Survey's operational footprint. The Data Acceptance Procedure is an internal document, but our remit is also defined for our users on the 'About Us' page (https://www.bas.ac.uk/team/business-teams/information-services/uk-polar-data-centre/) on our external website and linked from the 'depositing data workflow' (https://www.bas.ac.uk/data/uk-pdc/data-deposit/) for ease.

Once the data have been assessed against the NERC Data Value Checklist and the Data Acceptance Procedure there are three possible outcomes: a.) the data are not considered suitable for deposit; b.) the data are not considered suitable for deposit with the PDC but we can recommend an alternative repository for the dataset (another NERC Data Centre or more specialist subject specific repository, for example); or c.) the data are suitable to deposit with the PDC.

All suitable deposits must provide metadata for each dataset before it will be accepted by the data centre. The PDC has metadata guidance available on our website (https://www.bas.ac.uk/data/uk-pdc/metadata-guidance/), which includes a metadata template for depositors to complete. If the metadata provided is insufficient, the assessor will liaise with the depositor to complete and resolve any queries.

Metadata are made available via the PDC's metadata catalogue, the Discovery Metadata System (DMS) (https://data.bas.ac.uk/). The DMS complies with the ISO 1939 XML implementation schema for ISO 19115 Geographic Information. All metadata used to issue a DOI is automatically validated against the DataCite Metadata Schema 4.3 to ensure compliance. The automatic validation code is updated as new Datacite schema versions are released. Any issues identified by the validation process are resolved in liaison with the depositor.

Data are checked against a Data Quality Checklist (https://www.bas.ac.uk/wp-content/uploads/2020/04/PDC_Data_Quality_Checklist.pdf), which includes checks on file names, variable names and data formats. This check is recorded on the Data Accessions Tracking Log. Given the breadth of data the PDC hold, there are no strict data format requirements but data should be provided in an open, interoperable format where possible (see section fourteen); this is explained in the format guidance (https://www.bas.ac.uk/wp-content/uploads/2020/06/PDC_Data_Format_Guidance.pdf). Proprietary formats are considered if there is no open alternative, or if conversion would result in significant data loss. These are dealt with on a case by case basis in consultation with the depositor.

Once data have been accepted by the data centre they are typically kept in perpetuity, given their long-term value and irreplaceability. Should a dataset be superseded then the removal of the original dataset from the publicly accessible collection is assessed on a case by case basis. Should a dataset be removed, the metadata will remain available and the data will be available on request. It is made clear in the metadata that the data has been superseded and the updated version is linked to from the superseded version.

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

Response:

The PDC follows the best practice guidance as described in our digital data preservation policy (https://www.bas.ac.uk/wp-content/uploads/2020/05/PDC_digital_data_preservation_policy_v1.0.pdf). This describes how our processes and procedures are based on the Open Archival Information System (OAIS) reference model.

The PDC has a set of internal procedures that outline how data are handled throughout the accession process from the point of initial contact by the depositor to the publishing of the dataset and issuing of a DOI. These procedures include: a data accession procedure, a data acceptance procedure and an internal workflow. These are stored in a secure password protected storage area with access for PDC staff only. Relevant information is available publicly including a simplified version of the workflow (https://www.bas.ac.uk/data/uk-pdc/data-deposit/), a Data Transfer Agreement (https://www.bas.ac.uk/wp-content/uploads/2019/06/Data-Transfer-Agreement.docx) and guidance on providing quality metadata (https://www.bas.ac.uk/data/uk-pdc/metadata-guidance/). The process of issuing a DOI is documented in a structured Gitlab issue board system. Before data are published, a second data manager will check that all the procedures

have been completed, documentation is complete and that the data are ready for publishing. This check is recorded in the Accessions Tracking Log. PDC procedures are under constant review and a non-conformance log is in place to document any divergence from PDC procedures so that we can learn from mistakes and improve procedures/staff training when necessary.

The PDC has a clear understanding where all the data it accessions are stored. There is a specific area for data that have been issued with a DOI, as these need to be kept permanently. Embargoed datasets are kept separately from those publicly available and moved to the area for publicly available data when the embargo is over. The NERC data policy guidance stipulates that a reasonable embargo period is a maximum of two years from the end of data collection. We will release datasets from their embargo within this period unless the depositor provides an acceptable justification for a longer period of embargo. In these exceptional cases a longer embargo period is agreed and the data are released at the agreed date. BAS long-term datasets managed by the PDC are held in other known storage areas or in relational databases. These areas are all secure, with password access (see section sixteen) limited to the PDC. There are other BAS datasets which are held by BAS scientists and not yet accessioned into the PDC. Work is ongoing to increase the flow of data from scientists into the PDC so that they can be documented, managed for the long-term and made available for reuse.

The data storage system is backed up nightly, with any differences since the previous backup copied to a separate disk and tape storage system kept in a different fire zone. A copy of BAS's data is kept offsite on digital tapes, stored by a commercial data storage company in secure facilities. Once a week the offsite copy is updated with the changes since the previous week. During 2020, the backups will be migrated to a disk-based system, with an offsite system at the British Geological Survey Keyworth site. This offsite copy will be updated nightly over the internet, reducing the chance of data loss from a significant disaster at the Cambridge site and allow for a quicker restore than is currently possible from the tape offsite copy. BAS IT undertake regular checks on the physical media used and replace tapes as necessary to prevent issues with degradation. All physical media are being phased out, so that all storage will be on disk. There is an active programme in place to replace these on a four yearly cycle. All systems have ongoing scans for viruses and malware with alerts sent to the IT Service desk if any activity is detected so that it can be investigated.

To further improve our practice, the PDC is in the process of implementing automated scripts which will run regularly to validate checksums and ensure there have been no changes to the data held by the repository. A test system has been developed which runs a script daily to check the published data area for any changes to which files are present and their checksums. Any changes are logged and a warning message is sent to enable further investigation. This system will be fully implemented within the first half of 2021. File formats are recorded in Gitlab when the DOI for a dataset is minted allowing the PDC to keep track of the different file formats that are held. Our preference for open and predominantly ASCII-based formats helps to reduce the risk of files becoming unreadable but we monitor the types of format that are held and will put in place a migration plan whenever there is risk that a format might become unreadable in the future.

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

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The PDCs role as the national polar and cryosphere data centre assumes an indefinite retention of the data in its care. The NERC retention schedule stipulates that data assessed of long-term value should be kept permanently and data which has been given a DOI (an increasing proportion of data held by the PDC) will be kept in perpetuity. Most data held by the data centre are unique, irreplaceable and of long-term value.

As outlined in section nine, the PDC has a digital data preservation policy (https://www.bas.ac.uk/wp-content/uploads/2020/05/PDC_digital_data_preservation_policy_v1.0.pdf) describing how the data centre ensures the longevity of its digital information assets. It includes the model under which we operate, the legal and regulatory policies that apply, and addresses factors which risk making the data unusable and/or inaccessible. Preservation plans are also specified on a dataset level for all NERC-grant funded data in the respective project Data Management Plan.

For all datasets submitted to the PDC, the responsibility for the data is transferred from the depositor to the PDC upon deposit. A Data Transfer Agreement (https://www.bas.ac.uk/wp-content/uploads/2019/06/Data-Transfer-Agreement.docx) is drawn up between the data centre and the depositor, describing each party's responsibilities. It allows, for instance, the PDC to, "preserve the integrity of the data and protect it from loss or damage (e.g. carry out regular secure back-ups, format migration etc.)" but also to "copy, transform, and store the data, as well as provide access to them".

Standards accepted by the data centre are described in the Preservation Plan. The metadata template (https://www.bas.ac.uk/wp-content/uploads/2018/12/PDC_Metadata_Guidance_version7.docx) offers further guidance to depositors on the standards that the PDC works to and the information they are required to provide at a dataset level. The depositor is required to document details about the dataset (title, file size, name, format etc.), policies and legislation affecting the dataset, details of supporting documentation that will be submitted alongside the data, details on availability and access (embargo period, creation of web map services), and details on licensing, IPR and copyright.

Data are accepted by the data centre only if the Data Transfer Agreement has been formally agreed, the metadata has been provided as requested and the data are provided in the agreed format. Once the Data Transfer Agreement has been signed, the UK PDC assumes responsibility for the data resource.

To reduce the risk of obsolescence, files are generally accepted in a non-proprietary format; as explained in the PDC's format guidance (https://www.bas.ac.uk/wp-content/uploads/2020/06/PDC_Data_Format_Guidance.pdf). In some specialised scientific areas, proprietary formats that are used extensively by the relevant community and for which there is no equivalent non-proprietary format, may be accepted. Each dataset within the preservation system follows a consistent directory structure for storage. The PDC currently migrates data formats as and when required. We are developing a process in the first half of 2021 that will record the format of files (in addition to its checksum, see section nine) and file formats are also recorded in GitLab when DOIs are minted. This knowledge of formats enables us to be confident that we know precisely what data formats we hold and allow us to develop a migration strategy for particular formats when necessary. A formal migration workflow will be developed during 2021 and added to the Digital Preservation Policy.

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

Response:

All data stored by the PDC are required to have comprehensive metadata to allow discoverability and re-use. Data and metadata are checked by PDC staff to ensure completeness using the Data Quality Checklist (https://www.bas.ac.uk/wp-content/uploads/2020/04/PDC_Data_Quality_Checklist.pdf) which forms part of the internal data accessions procedure. Checks include that the data are well described, file and variable names are meaningful, and data values contain no obvious errors. A record that these check have taken place is recorded in the Accessions Tracking Log. 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. The PDC specialist data managers have relevant domain knowledge enabling them to understand the contextual metadata, ask relevant questions when information from the depositor is lacking and ensure sufficient information has been received to allow the data deposited to be reused by others. The PDC requires that the data supplied has a signed Data Transfer Agreement whereby the depositor has agreed the data "are of a good technical quality, suitable for archiving and sharing" and "to supply sufficient metadata to enable discovery and re-use of the data". The depositors are encouraged to share information on the data quality including any quality control procedures followed, the accuracy of the data, factors affecting the data, and any loss of data. Any information provided to the data centre on data quality is included in the metadata record for the dataset. These details are in the 'Additional Information' section of the metadata record under lineage and quality (e.g. https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/01318). Older catalogue records do not always have this information included but we have recently reviewed all catalogue records and, during 2021, we will improve the detail and quality of older records where possible.

BAS data that are processed by PDC staff contain quality flags where appropriate. Typically this will show three levels of quality assurance with one indicating the data have been through a quality control process and that no issues have been flagged. Higher levels of flag indicate that an issue has been found with the quality of the data. Details of issues found are kept with the data. Only data with a quality assurance of one are data are served out publicly by default. For example, the interface to BAS (http://basmet.nerc-bas.ac.uk/sos/) only provides access to data which has been quality controlled by the met team, and when multibeam data are requested, the PDC will always provide the best quality data available for the area requested. Biological species names are checked against relevant taxonomic indexes e.g. WoRMS for marine species.

Metadata are made available via the PDC's metadata catalogue, the DMS (https://data.bas.ac.uk). Records in the

catalogue comply with the ISO 19115 standard and, for datasets with persistent identifiers, metadata also complies with the DataCite Metadata Schema. The technical standards we comply with are recognised within the scientific data management community and a CSW endpoint (https://docs.api.bas.ac.uk/services/data/metadata/csw/v2/usage/) allows our metadata to be harvested by other catalogues such as the NERC Data Catalogue. The discovery metadata are also evaluated by other peers through an annual review conducted by the metadata subgroup of the NERC Data Operations Group.

Data users have the option to comment on data and/or metadata supplied by the PDC via completion of our contact form accessed via the dataset, or by directly emailing the data centre (https://www.bas.ac.uk/data/uk-pdc/contacts/).

The metadata records in the DMS contain full citation information for those needing to acknowledge their data source. In addition, links to associated data, publications and project information are added to the metadata where applicable, to support interpretation of data deposited.

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

Reviewer 2

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

Response:

The PDC has documented policies and procedures in place to cover the data lifecycle from the pre-ingestion to the preservation phase. These processes are detailed on the BAS website, along with clear guidance to depositors and users about handling of data. Each step in the workflow is recorded in our Accessions Tracking Log so that we have a permanent record that all steps have been undertaken.

All data deposited to the PDC needs to fit in our remit (https://www.bas.ac.uk/team/business-teams/information-services/uk-polar-data-centre/). The PDC will assign a data manager to each deposit request, who will initially assess the data using the NERC Data Value Checklist (https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/) and the internal Data Acceptance Procedure. They will then liaise with the depositor to discuss key details of the deposit and data transfer methods.

If the data are suitable for ingestion, the depositor needs to provide:
• A signed Data Transfer Agreement (https://www.bas.ac.uk/wp-content/uploads/2019/06/Data-Transfer-Agreement.docx) that sets out the PDC's and the Depositor's responsibilities with regards to transferring, sharing and distributing data.
• A metadata questionnaire (https://www.bas.ac.uk/wp-content/uploads/2018/12/PDC_Metadata_Guidance_version7.docx). Extended information and guidance notes are provided on BAS website to help depositors fill in the questionnaire (https://www.bas.ac.uk/data/uk-pdc/metadata-guidance/).
• Data in the required format and following the PDC's Data Quality Checklist (https://www.bas.ac.uk/wp-content/uploads/2020/04/PDC_Data_Quality_Checklist.pdf).

The assigned data manager (who specialises in that data type or domain) will check both data and metadata to make sure data are fit for publication and re-use, and liaise with the depositor to gather further information where necessary. From this stage, data are stored securely in an unpublished repository that only the PDC team can access. This protects incomplete data accessions and embargoed data from being exposed prior to the planned date.

Each data deposit is assigned an accession deposit ID and documented internally within an accession database, enabling any member of the PDC to follow the stage of each deposit. The data manager creates a metadata record on the DMS (https://data.bas.ac.uk/). When the data are ready to be published, data are distributed through a RAMADDA data repository and linked to the metadata record. If an embargo has been requested, the data manager will update the metadata record with the embargo details and create a post-dated task to lift the embargo at the agreed date.

For each deposit we assign a DOI as standard and as part of the DOI minting procedure, a DOI tracking system has been developed on the internal Gitlab system. As part of the procedure, another data manager quality assures the work carried out by checking all the required information has been provided to make sure the DOI DataCite metadata are well described.

A simplified workflow of our data ingestion procedure is described on our website for our users, with links to relevant documentation (https://www.bas.ac.uk/data/uk-pdc/data-deposit/). The PDC have a more detailed internal accessions procedure to ensure our data managers are all working consistently and have the guidance to make necessary decisions

throughout the process. Changes can be requested by any member of the data centre but must be put forward to the rest of the team for review and finally approved by the Head of the PDC. The changes are implemented once the relevant documented procedures are updated and published. Deviations from the published workflow are recorded in a non-conformance log to enable us to improve our procedures and staff training.

Feedback is regularly requested from users on their interaction with us (https://www.bas.ac.uk/pdcfeedback), and should they suggest or request a change to our workflows, a metadata record or dataset, this will be reviewed by the data manager for that data type or service. Suggested changes to our workflow will be reviewed in the same way as suggested changes made by a team member. Requests to make minor changes to metadata are carried out by the data manager in liaison with the user. If a requested change would lead to a new DOI (i.e. a new version of a dataset), the relevant data manager would assess the files and get a second person to approve the changes following our usual workflow.

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

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The PDC maintains an online discovery metadata catalogue at https://data.bas.ac.uk. The catalogue provides controlled search options based on topic and location and a free text search on title, abstract and keywords. The catalogue was originally developed to be compliant to the Directory Interchange Format (DIF) and has been subsequently modified to comply with the ISO 19115 standard.

Discovery metadata records are made available for machine harvesting through a Catalogue Service for the Web (CSW) API (see https://docs.api.bas.ac.uk/services/data/metadata/csw/v2/ for details). Records provided through the CSW are harvested by the NERC Data Catalogue (https://data-search.nerc.ac.uk/) which is itself harvested by a number of national and international metadata aggregators. Records published by the UK PDC also appear within the Antarctic Master Directory, a subset of the NASA Global Change Master Directory (GCMD). The UK PDC is listed in the Registry of Research Data Repositories (http://doi.org/10.17616/R3C01K).

The PDC provides persistent identifiers for datasets through DataCite DOIs. The data centre has a documented process for issuing DOIs and all new accessioned datasets will be assigned one. In order to issue a DOI, the dataset must have been fully ingested into the data centre and be described with sufficient metadata to enable discovery and reuse. A dataset DOI issued by the UK PDC will direct a user to the relevant metadata record landing page within the online catalogue.

A recommended citation string is recorded for all datasets with a DOI and is usually provided along with the dataset within the RAMADDA data repository (e.g. https://ramadda.data.bas.ac.uk/repository/entry/show?entryid=37cafd85-e84f-416c-8c09-51a5ec9b001c). New catalogue records for which datasets have been issued a DOI, include a Get Data link that allows a user to directly download the dataset. Older catalogue records do not all have a Get Data link and in these cases a user had to request the data from the PDC. An initial response is typically made to the requester within two working days and data will normally be sent within seven days unless the request is particularly complex, the data are embargoed or not held within the PDC. Approximately 80% of catalogue records have a direct link to the data shown under the Get Data section of the catalogue record (see https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/01316 for an example) and this will increase over time now that it is normal practice to provide this for all data that we accession. Catalogue records are available for embargoed datasets and it is shown under the access constraints section of the record that the data are embargoed (https://data.bas.ac.uk/full-record.php?id=GB/NERC/BAS/PDC/01433). It would still be possible for someone to request these datasets and the request would be discussed with the lead author of the dataset.

The PDC is actively working on increasing its use of persistent identifiers. Currently DOI metadata includes ORCIDs for people and Research Organisation Registry identifiers for organisations. We actively use, and add to, the NERC Vocabulary Server which provides described standardised terms e.g. to refer to instrumentation or sampling method.

Improvements to aid discovery and identification of datasets published by the UK PDC are ongoing. Specific tasks include:
- Providing an improved online catalogue interface with additional faceted search options and better presentation of metadata content;
- Ensuring that the CSW contains as much information as the records in the underlying database – currently this is not the

case;

- Issuing dataset DOIs for all records within the PDC metadata catalogue – datasets accessioned prior to 2017 do not all have DOIs;

- Populating recommended citation strings for all records and making citation information visible on every metadata record landing page.

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:

3 – The repository is in the implementation phase

Response:

Discovery metadata, as described in section 13, are required when data are deposited in the PDC. 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, abstracts, keywords etc. This also enables datasets to be exposed through appropriate external gateways e.g. the NERC Data Catalogue Service and the Antarctic Master Directory. Provision of detailed discovery metadata enables users searching for data to make an initial, high-level assessment of whether the data resource being described is suitable for their requirements.

In addition to discovery metadata, the PDC captures as much supporting information as possible alongside the raw data. This supporting information is checked as part of the data ingestion process to ensure the data can be reused without recourse to the data authors. The PDC aims to ensure that the dataset is complete and as described and that the description is meaningful (e.g. units are included for all relevant variables, data collection techniques are described, instrumentation is described, acronyms and abbreviations used are defined). We encourage the use of internationally recognised vocabularies to describe instrumentation and variables. For example, marine events on BAS research vessels are described using terms from the NERC Vocabulary Server (https://www.bodc.ac.uk/resources/products/web_services/vocab/) which is used internationally to describe marine data sets. We regularly request the addition of new terms to this service. The NERC Data Centres are currently working together to expand this service to better cover domains outside of the marine realm. The metadata record also includes links to directly relevant scientific publications and field reports to further support understanding and reuse of the data.

The PDC provides guidance on the file formats accepted by the data centre to encourage the deposit of data in formats that are at less risk from technology/software obsolescence or require proprietary software to open (https://www.bas.ac.uk/wp-content/uploads/2020/06/PDC_Data_Format_Guidance.pdf). The preferred formats are generally non-proprietary, or open industry-standard formats that can be used by anyone accessing the data. This range of data types is dynamic and evolves over time as the type of science and technologies used in polar science changes.

The data that PDC accessions at the end of scientific projects are normally published in the same format in which they were received. However, BAS long-term datasets that are managed and/or processed by the PDC, maybe converted into other formats for long-term storage such as a relational database. This gives the PDC flexibility in how the data can be made accessible to others e.g. through a tailored search interface (http://basmet.nerc-bas.ac.uk/sos/) or via an API (https://www.bas.ac.uk/api/).

Our mitigation strategies for data format evolution are as follows:
1. Where possible use generic or simple data formats (e.g. csv) that are likely to be more stable
2. Encourage use of open formats that have community-driven backward compatibility (e.g. NetCDF).
3. Further develop our software library for converting data from proprietary formats to open formats. Share these tools with other NERC data centres and more widely if appropriate.
4. Monitor, and contribute to, development of data formats and conversion tools in relevant domains and expert communities and make use of these when relevant.

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:

3 – The repository is in the implementation phase

Response:

The PDC's data are stored on BAS's on-premise central data storage system, managed by the BAS IT department. The data storage system is kept in a purpose-built facility, with an Uninterruptable Power Supply to prevent uncontrolled loss of power and environmental control and monitoring (temperature and humidity) to keep these factors within the manufacturer's guidelines. Physical access to the facility is controlled and restricted to IT, Estates and Health & Safety staff. Individual servers and storage systems within BAS's IT infrastructure have redundant power supplies, network connections and hard disks that ensure a single failure of any one of those high-risk components does not cause loss of data or interruption to service.

The data storage system is backed up nightly, with any differences since the previous backup copied to a separate disk and tape storage system kept in a different fire zone. A copy of BAS's data is kept offsite on digital tapes, stored by a commercial data storage company in secure facilities. Once a week the offsite copy is updated with the changes since the previous week. The backup system is focused on providing a cost-effective recovery from common errors (accidental deletion, minor technical faults). The offsite copy allows recovery from major errors and serious incidents (such as a site wide disaster across the BAS offices). In 2020, the backups will be migrated to an off-site disk-based system based at the British Geological Survey Keyworth site. This will be updated nightly over the internet, reducing the chance of data loss

from a significant disaster at the Cambridge site and allowing a quicker restore.

BAS has high-speed access to the UK Joint Academic Network (JANET) which ensures the PDC can meet the data access needs of its designated community. There is dual redundancy in the BAS site JANET feed with auto failover, meaning that in the event of one connection failing, the internet feed will not be lost. The BAS network authentication also has fail over capability via servers hosted at a different NERC location. This allows continued connectivity to email in the event of a site failure. The research nature of the infrastructure does not permit 24/7 support for the system to ensure round the clock support for incoming data feeds and support of end-users. However, automated checks of systems allows for problems to be identified prompty and resolved in a timely manner.

Much of the generic software used by the PDC is managed, licenced and upgraded by the IT department. All network connected PCs are audited so IT staff are aware of what software is currently installed. The IT department records when software licences are due for renewal and there are plans to implement a full Software Asset Management program providing resources for this is available. BAS IT will be developing an IT continuity plan in 2021 to clearly articulate the strategy in place for this. There is also a five year plan in place to obtain ISO 27001 certification. Currently PDC do not have an SLA in place with the BAS IT team, but we will work together to develop one in 2021.

PDC staff expertise includes software development for more bespoke applications and these developers ensure that these applications are kept up-to-date and maintain the functionality required. Data that have been assigned a DOI are managed using an instance of the open source software, RAMADDA (https://ramadda.org/) to provide a web interface for data access. Metadata records comply with the ISO 19115 standard and, for datasets with persistent identifiers, metadata also complies with the DataCite Metadata Schema.

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

Response:

PDC's data are kept secure by BAS's underlying IT infrastructure. BAS is committed to taking all necessary precautions to ensure the physical safety and security of all data resources it preserves and ensures compliance with the NERC Information Security Policy. BAS has systems in place to ensure only authorised people are in the building. In addition, the computer server rooms are equipped with a security-protected card access system with access restricted to relevant people. As outlined in the section above, systems are in place to ensure that recovery is possible in the event of either a systems failure or a major disaster involving the BAS site. Access to data storage areas, databases and other software is managed using file system permissions and usernames and passwords. The data storage system only allows access to specific internal BAS systems; those managed by BAS IT or delegated groups. Write access to PDC data storage is restricted to PDC staff, and read access is controlled with user and group permissions.

Access to the BAS network is controlled by a site-wide firewall to restrict access to specific known systems. A secure VPN allows BAS staff remote access. BAS has Cyber Essentials accreditation (https://www.ncsc.gov.uk/cyberessentials/overview) and all staff are required to undertake training in cybersecurity provided by Cybsafe (https://www.cybsafe.com/). There is a strategic plan to achieve ISO 27001 (Information Security Management System) compliance by 2025. Best practice of ISO 27001 is implemented in a majority of areas, e.g. password policies (attached for reviewers), file security. BAS adheres to all UKRI policies and procedures for security. NERC operate an Information Security Assurance Group (ISAG) including members from BAS IT which meets regularly. There is a NERC system for reporting security incidents, which are regularly reviewed by ISAG and UKRI. NERC's Information Security Incident/Data Breach Response Procedure is not publicly available but attached for reviewers.

The PDC has a process in place to ensure that data held under embargo are not available to the public until the agreed end of the embargo period. Data held under embargo are stored in the PDC secure data storage area where they remain inaccessible to the public. Embargo periods are monitored and when an embargo period has expired the data are made publicly available and accessible through the data catalogue (see section 12).

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:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: