UK Polar Data Centre Digital Data Preservation Policy

Version 2 released: 12 December 2023

1. **Purpose**

This policy documents the data preservation policy of the UK Polar Data Centre (PDC), detailing its scope, objectives and drivers. The PDC is a Natural Environment Research Council (NERC[1]) data centre hosted by the British Antarctic Survey (BAS). The aim is to ensure the longevity of the digital information assets held by the PDC in a sustainable way by addressing the factors which risk making them unusable and/or inaccessible.

2. **Mandate**

The PDC is the focal point for Arctic and Antarctic environmental data management in the UK[2]. We coordinate the preservation and management of polar data from UK-funded research and support researchers in complying with national and international data legislation and policy such as the NERC data policy[3] and Article III 1(c) of the Antarctic Treaty[4]. It is the UK's National Antarctic Data Centre.

3. **Scope**

The scope of this policy is limited to the PDC's data collections. It covers all research data and associated project management information for which PDC has taken on responsibility. This may be through an official deposit or where the PDC captures the data directly from instrumentation or user entry. This includes:

- NERC (or BAS) owned data
- Data generated by NERC funded grants
- Third party data (data we hold on an in-licensed basis)
- Data generated by non-NERC funded grants that is deemed to be of significant long-term value to the polar scientific community

It excludes material held within the BAS Archives and research papers which are stored in the NERC Open Research Archive (NORA).

4. **Objectives**

The primary aim of the PDC is to act as the NERC Data Centre for the UK polar and cryospheric science community. Our remit and a description of the types of data we take is defined in the PDC data acceptance procedure[5]. The PDC takes in data resources and assumes responsibility for the long-term preservation and accessibility of data in digital

---

[1] **NERC** is part of UK Research and Innovation www.ukri.org
[2] https://www.bas.ac.uk/team/business-teams/information-services/uk-polar-data-centre/
[3] https://www.ukri.org/wp-content/uploads/2022/03/NERC-080322-policy-data-021219.pdf
[4] https://www.ats.aq/e/antarctictreaty.html
[5] https://www.bas.ac.uk/wp-content/uploads/2023/12/PDC_Data_Acceptance_Procedure_v2.pdf

form. This policy outlines the key actions and rationale behind the actions taken to ensure that the data held by the PDC are permanently accessible in a form that is fit for purpose for all the end users of the service provided.

The specific aims of the preservation policy are to:

- provide reliable instances of data resources to the designated user community
- maintain the integrity and quality of the data resources
- ensure all data resources are protected
- ensure the relevant level of information security is applied to each data resource
- instil good practice in active preservation management

5. Requirements

A core function of the PDC is to acquire, maintain and manage related digital resources of value to the polar and cryospheric science community, and to promote and disseminate these resources as widely and effectively as possible. Therefore, the PDC has developed a series of requirements which it strives to ensure are followed as closely as possible to meet the needs of its designated community:

- the data resources it accessions are accompanied by sufficient documentation to enable their re-use;
- the data resources are checked and validated according to strict data and documentation ingestion procedures;
- the data resources are catalogued according to appropriate metadata standards;
- the data resources, documentation and metadata are kept in conditions suitable for long-term archival storage;
- the authenticity, integrity and reliability of data resources preserved for future use are retained;
- the data resources are held with rights for control to enable changes in format or structure to ensure long-term preservation;
- the basic preservation actions undertaken by the PDC are of a uniformly high standard regardless of the perceived value of any data resource.

6. Legal and regulatory framework and other policies

There are numerous legal and regulatory policies that impact on the management of data held by the PDC. This policy helps the PDC meet its legislative and accountability requirements and the expectations of its user community. The PDC must have legal rights to preserve any digital content kept in its archives and will not ingest materials that have unclear ownership or unresolved rights issues.

In preserving its data collections, the PDC follows:

- Data Protection Act 2018 (DPA 2018)[6]

---

[6] https://www.legislation.gov.uk/ukpga/2018/12/pdfs/ukpga_20180012_en.pdf

- General Data Protection Regulation (GDPR)[7]
- Freedom of Information Act 2000[8]
- Environmental Information Regulations 2004[9]
- Public Records Act 1958[10] and 1967[11]
- English/UK law for commercial agreements and contract law
- The Antarctic Treaty[12]

Other policies with an impact on digital preservation include:

- NERC Data Policy[13]
- UKRI Information Management Policy[14]
- NERC Good Research Conduct and Research Integrity Statement[15]
- UKRI Data Protection Policy[16]
- BAS science strategy[17] and business plan
- PDC standard operating procedures

The policy is also informed by:

- the aims and objectives of the PDC
- NERC's strategic and operational plans
- the needs of the users of the PDC
- best practice promoted by the Standing Committee for Antarctic Data Management and/or the Arctic Data Committee
- archival theory and practice
- relevant national and international frameworks

The PDC will follow the broad guidance given in standards and best practice guidance to support the level of preservation required.

These include:

- Core Trust Seal data repository requirements
- ISO 19115-2 (Geographic Information - Metadata - Part 2: Extensions for Imagery and Gridded Data) (ISO 19115-2:2009)
- DataCite Metadata Schema (4.3)

[7] https://gdpr-info.eu/
[8] https://www.legislation.gov.uk/ukpga/2000/36/contents
[9] https://www.legislation.gov.uk/uksi/2004/3391/contents/made
[10] https://www.legislation.gov.uk/ukpga/Eliz2/6-7/51
[11] https://www.legislation.gov.uk/ukpga/1967/44
[12] https://www.ats.aq/e/antarctictreaty.html
[13] https://www.ukri.org/about-us/nerc/our-policies-and-standards/nerc-data-policy/
[14] https://www.ukri.org/publications/ukri-information-management-policy/
[15] https://www.ukri.org/publications/nerc-good-research-conduct-and-research-integrity-statement/
[16] https://www.ukri.org/publications/data-protection-policy/
[17] https://www.bas.ac.uk/science/our-research/our-strategy/

We also implement these metadata profiles (for ISO 19115):

- INSPIRE via the INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (1.3)

7. Roles and responsibilities

Staff at the PDC are employed by the British Antarctic Survey (BAS). Accountability pertaining to preservation and re-use falls to:

- **BAS Director of Innovation and Impact** - Business owner of the PDC Digital Preservation function.
- **Head of the PDC** - Strategic alignment of digital preservation, optimisation of available resources.
- **PDC staff** - Expertise in digital preservation, covering the development and implementation of digital preservation policy, strategy and associated workflows. Ingest, store, help deliver and preserve the data and provide guidance to users.
- **Data creators and depositors (internally and externally)** - Appropriate actions to safeguard the data they create and deposit, including the use of appropriate file formats and provision of descriptive and technical metadata.
- **IT Systems and Network Support** - Ensure infrastructure used for storing digital records and materials is fit for purpose. Apply general data security.

8. Model

The PDC broadly follows the guidance provided in the Open Archival Information System (OAIS) reference model in its approach to long term preservation. The stages involved in curation of digital assets are: (equivalent OAIS functional entities in parentheses)

Identification	(Administration)
Preparation	(Administration)
Data transfer	(Ingest, Archival storage, Data management, Administration)
Access	(Access, Administration)
Administration/Housekeeping	(Data management, Administration)
Data centre management	(Preservation planning)

The PDC has established a set of defined processes and guidance to manage all stages. An issue tracking system, organised file systems and discovery metadata catalogue are used to track and document all work.

8.1. Identification

Data identification is used to quickly determine whether or not the data resource is suitable for ingestion by ascertaining:

- If the data are within the scope of the PDC's remit

- If the data are of sufficient quality
- If the data are documented sufficiently to aid reusability

Depositors are provided with guidance as to what is expected from them including:

- Accepted file formats and file naming conventions
- Accepted supporting documentation
- Metadata required

The identification stage helps to improve data quality, comprehensibility and accessibility by enforcing minimum standards of quality at the point of deposit. They ultimately help reduce the time required for the subsequent ingestion stages as data are submitted at a standard which requires less processing.

8.2. Preparation

The preparation stage involves negotiating service agreements (submission agreements) with the depositor. This includes them signing a Data Transfer Agreement[18] and agreeing any embargo periods for the data.

8.3. Data transfer

Data transfer involves the receipt of data and associated metadata from a depositor. At this stage data are examined to ensure they are uncorrupted, complete and what they purport to be. Data and metadata are stored in an appropriate secure place on the BAS Central Storage System.

Discovery metadata are finalised and published in the PDC data catalogue which facilitates discovery and access to the data. A Digital Object Identifier (DOI) is then assigned to a data resource. The published metadata are also harvested by additional catalogues, such as the NERC data service[19] and Antarctic Metadata Directory[20], further promoting the deposited data resources.

8.4. Access

The access process makes the data resource available to download. The PDC does not alter data once all the data quality checks have been undertaken and any revisions needed completed to the satisfaction of both the depositor and the PDC.

Usually, the access process follows immediately from the data transfer process. However, data may be embargoed (subject to the terms agreed with funders) in which case there can be a delay of several months/years between data transfer and access. Embargo periods are monitored, and data are made public once the agreed embargo period has passed, or earlier if requested by the depositor.

---

[18] https://www.bas.ac.uk/wp-content/uploads/2019/06/PDC_Data-Transfer-Agreement.docx
[19] https://data-search.nerc.ac.uk/
[20] https://search.earthdata.nasa.gov/portal/amd/search

8.5. Administration and housekeeping

Day to day administration of the PDC ensures that metadata and data are continually monitored and maintained.

- **Data integrity** – The checksum of data that are published by PDC is recorded and monitored daily. Any change to the data triggers an alert which is then investigated and, when necessary, rectified by data centre staff.
- **Version control/change procedures** - Data ingested into the PDC will normally be given a DOI which precludes any alterations/updates being made to the data. If a dataset is later found to be in error, a new version must be deposited with the PDC. Online access to the erroneous version will be rescinded but it remains available on request. If an updated version of a dataset is received, this will be published with a new DOI. Both versions will remain available, but the earlier version will have a clear link to show that an updated version is available.
- **Review procedure** – PDC undertakes a number of scheduled proactive reviews such as regular review of policies and procedures (at least annually) and general data/metadata quality control spot checks on a random subset of records (biannual). In addition, feedback from users or changes in standards/data formats lead to ad hoc reviews. These reviews can lead to a widespread review of all published metadata/datasets if appropriate. Decisions on when this is needed are made at the regular PDC meetings.
- **Withdrawal** -The PDC, in line with NERC, has a minimum retention period of ten years after completion of the research, after which data are periodically reviewed and potentially discarded. However, data which has been given a DOI (the majority of data held by the PDC) will be kept in perpetuity.

8.6. Data centre management

The data centre is managed using a series of defined, documented procedures.

The PDC meets at least monthly to deal with problems arising in day-to-day running of the data centre, plan future development priorities, deal with non-conformance issues and any other issues related to data centre management.

The Head of the PDC and BAS Director of Innovation and Impact are responsible for ensuring that the PDC is resourced appropriately.

9. Preservation planning

9.1. Strategy

The preservation strategy of the PDC aims to maintain a flexible preservation system that can evolve to meet the demands of changing technology and developing user expectations. The PDC has chosen to implement a preservation strategy based upon open and available file formats wherever possible. The same ingestion procedure is

used for all data resources and no judgement is made on the scholarly value of the datasets once they have been identified as suitable for deposit with the PDC. All datasets accepted for deposit must be accompanied by supporting documentation of sufficient quality to enable re-use over the long-term.

To reduce the risk of obsolescence, files are generally accepted in a non-proprietary format. 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.

Our mitigation strategies for data format evolution are as follows:

- Where possible use generic or simple data formats (e.g., csv) that are likely to be more stable.
- Encourage use of open formats that have community-driven backward compatibility (e.g. NetCDF).
- Maintain a suite of software tools for converting data from proprietary formats to open formats. Monitor, and contribute to, development of data formats and conversion tools in relevant domains and expert communities and make use of these when relevant.

9.2.   IT infrastructure and security

The preservation of the PDC's data relies on a BAS IT infrastructure that is fit for purpose and is continually monitored and periodically reviewed to ensure timely upgrades in both hardware and software.

The PDC's data are stored on BAS's on-premises 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 storage system kept in a different fire zone.

A copy of BAS's data is kept offsite on another disk storage system at a different NERC Centre, the British Geological Survey, at its Keyworth site. The offsite copy is updated nightly with the changes since the previous night. 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).

9.3.   Network security

BAS is committed to taking all necessary precautions to ensure the physical safety and security of all data resources it preserves. The repository rooms are equipped with a security-protected card access system.

All data are available on network shares to internal BAS systems. Access control is managed with file system permissions. Write access to PDC data storage is restricted to PDC staff, and read access is controlled with user and group permissions.

The data storage system only allows access to specific internal BAS systems; those managed by BAS IT or delegated groups (such as the PDC).

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, other access is provided via PDC maintained systems.

9.4.   Cooperation

The PDC has established productive working relationships with the other four NERC data centres (British Oceanographic Data Centre, Centre for Environmental Data Analysis, National Geoscience Data Centre and Environmental Information Data Centre), including liaison in determining the most appropriate data centre for resource curation. It also liaises regularly with other national Arctic and Antarctic Data Centres and other international domain specific repositories.

9.5.   Funding and resource planning

The PDC is dependent on funding from NERC to carry out its activities. The PDC is commissioned as part of the NERC Environmental Data Service[21] on a five yearly basis with the current funding awarded from April 2023-March 2028 with additional funds from NERC grants for the accessioning of new data.

Resource management for preservation of digital resources includes:

- technical infrastructure, including equipment purchases, maintenance and upgrades, software/hardware obsolescence monitoring, network connectivity etc.
- financial plan, including strategy and financing the PDC and commitment to long-term funding
- staffing infrastructure, including recruitment, induction and ongoing staff training

The PDC makes every effort to remain up to date with any relevant technological advances to ensure continued access to its collection. The PDC also implements a

---

[21] https://eds.ukri.org

programme of continual improvement in its procedures and how users interact with the data centre.

10. Policy Monitoring and Review

The preservation policy of the PDC is monitored and reviewed in light of changing technologies on an annual basis to ensure timely updates. The Head of the PDC initiates the review process in association with PDC staff. Implementation of the policy is monitored via the PDC's non-conformance process. This policy is available to all staff and members of the public via the PDC pages on the BAS website. Queries concerning the preservation policy should be directed to PDCServiceDesk@bas.ac.uk.