



Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: Environmental Information Data Centre
Website: <http://eidc.ceh.ac.uk/>
Certification Date: 16 August 2019

This repository is owned by: **Natural Environment Research Council (NERC)**



Environmental Information 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:

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

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

The Environmental Information Data Centre (EIDC) is the UK national data centre for terrestrial and freshwater sciences. It is hosted by the Centre for Ecology and Hydrology. The EIDC is funded via its parent organisation the Natural Environment Research Council (NERC).

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

The EIDC accepts environmental data in re-usable formats for curation, including long-term storage and public dissemination. Priority is given to data generated as a result of funding by NERC, however, data generated through other funding sources is also considered for deposit at the discretion of the EIDC.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

The designated community for the EIDC consists of a wide range of users of terrestrial and freshwater environmental data. Users are typically in academia, local authority organisations, government departments and agencies (e.g. Defra, Environment Agency, Natural England, Agri-food and Biosciences institute) industry and environmental consultants.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

A. Content distributed as deposited, 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

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

The EIDC aims to preserve and disseminate its data holdings for the long-term. The data held by the EIDC 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 EIDC facilitates scientific peer-review processes, open access and re-use of nationally important datasets and information products.

The EIDC has been in existence since 2000, when its host organisation, the Centre for Ecology and Hydrology (CEH) was formed. Four institutions: the Institute of Hydrology, the Institute of Terrestrial Ecology, the Institute of Freshwater Ecology and the Institute of Virology and Environmental Microbiology merged to create CEH. The data held by the data centre has been derived from more than 75 years of terrestrial and freshwater research programmes carried out at these and other institutions.

The EIDC has its own website (<http://eidc.ceh.ac.uk/>) which provides users with information on how to find, deposit & cite data, what support is available and our guiding principles such as our acquisition and preservation policies.

Depositors are required to provide detailed metadata and supporting documentation alongside the datasets offered to the data centre. Quality assessment procedures are carried out at every stage of the ingestion process. Supporting documentation of ingested datasets may be reviewed, updated and improved by data centre staff at any time.

Datasets deposited to the EIDC are not subjected to any changes during their lifetime at the data centre and are therefore distributed to data consumers just as they were when deposited. The exception to this might be where a user has spatially 'clipped' the data, in which case they would receive a sub-set of the data originally deposited, or the data may be re-projected at a user's request.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Other Relevant Information.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

ORGANIZATIONAL INFRASTRUCTURE

I. 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 Environmental Information Data Centre (EIDC) is a NERC Data Centre hosted by the Centre for Ecology & Hydrology (CEH). We manage nationally-important datasets concerned with the terrestrial and freshwater sciences. The EIDC is required to hold scientific programme outputs from NERC funded research and our mission is stated on our website (<http://eidc.ceh.ac.uk>).

We recognise the value of scientific data and the importance attached to the long-term professional management and preservation of data assets. Data is vitally important both as an evidence base for existing scientific projects and for future re-use. This belief is encapsulated in our Acquisition Policy (<http://eidc.ceh.ac.uk/policies/acquisition>) along with the NERC strategy (<http://www.nerc.ac.uk/about/whatwedo/strategy>) and Data Policy (<https://nerc.ukri.org/research/sites/data/policy/>).

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

II. 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

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

Licensing information, terms of use and any required citations/acknowledgements are included in the metadata record describing each dataset/service (e.g. <https://doi.org/10.5285/55bc4927-5d9b-4e64-b30e-f4f97c84b87c>).

A variety of different licences are in use but the majority of datasets are made available under the terms of the UK Open Government Licence (<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>).

The full range of licences in use are available at

<http://eidc.ceh.ac.uk/administration-folder/tools/ceh-standard-licence-texts>. The licences are managed in collaboration with the CEH Data Licencing Team. Many of the licences are kept for historic reasons - they apply to datasets ingested in the past but are no longer used for 'new' datasets.

Digital data licences include a termination clause in case of non-compliance with the stipulate terms and conditions, however, we currently have no means of identifying non-compliance unless it is brought to our attention by others.

A very small number of information products we hold incur a data preparation and provision charge. These are products that are generated from licensed third-party material for which we are obliged to pay a fee. The level of charges depends on the intended use of the data (e.g. whether commercial or academic). Any charges for data comply with the NERC Data Licencing and Charging Policy (<https://nerc.ukri.org/research/sites/data/policy/nerc-licensing-charging-policy/>) and in accordance with UK Government legislation. All such requests are currently dealt with on a case-by-case basis by the CEH Data Licencing team.

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 agreement.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

III. 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 EIDC is currently implemented as an integral part of the CEH infrastructure. A detailed formal continuity plan has been developed by CEH Computing Support (CCS). This is an internal document and is not available externally. The EIDC datasets are held within and accessible from the CEH infrastructure and as a result their future continuity is determined by that of the organisation. To comply with best practice and industry standards, data assets are backed up and archived according to stated policies that are implemented by the CCS team. For added security these backups are stored securely off-site.

NERC/UKRI manage the JASMIN super-computing centre at Harwell which currently hosts the atmospheric sciences data centre (CEDA) for NERC and NCAS (National Centre for Atmospheric Science). The EIDC (and the other NERC data centres) as part of their long-term strategy will be implementing a move to a cloud-based solution for data storage and delivery on the JASMIN infrastructure to provide a single data service for NERC data centres. Moving to this solution will provide long-term stability and also enable the EIDC to be more proactive in planning for rapid changes in storage and delivery methods.

All data held by the EIDC 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/#iso-19139). It is accessible and searchable in a metadata catalogue (<https://catalogue.ceh.ac.uk/eidc/documents>) and also harvested by metadata aggregator sites such as the NERC Data catalogue Service, Data.Gov UK and EU JRC.

A digital preservation strategy for the data centre is available publicly through the EIDC website (<http://eidc.ceh.ac.uk/policies/digitalPreservation>). This policy aims to encapsulate best practice for the UK and European community and describes the reference model that the EIDC has adopted. Preservation planning, IT architecture and funding and resource planning are also documented in the EIDC's digital preservation policy.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. 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

Reviewer 2

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

Response:

The EIDC 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). All staff working in the EIDC are required to undergo annual training on GDPR and Information Security.

The EIDC also complies with NERC Ethics Policy (<https://nerc.ukri.org/about/policy/policies/nerc-ethics-policy/>). This 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 also publishes a Research Grants and Fellowship Handbook

(<http://www.nerc.ac.uk/funding/application/howtoapply/forms/grantshandbook/>) that includes guidance on research ethics. Researchers are required to comply with the 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/>), which should be read in conjunction with the UK Concordat to Support Research Integrity (<https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/research-concordat.aspx>) and the guidance on Good Research Conduct and Research Integrity (<http://www.nerc.ac.uk/about/policy/policies/research-integrity/>). These policies and guidelines apply equally to researchers, support staff, research administrators, Research Council staff and all individuals contributing to the Research Councils' peer review process.

The EIDC 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 (<http://www.nerc.ac.uk/about/policy/foi/records-management-policy.pdf>) 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 EIDC has a rigorous and comprehensive ingestion process in place (<http://eidc.ceh.ac.uk/deposit>). Depositors are asked to indicate which policies and legislation apply to the dataset they are depositing in a Service Agreement to which consent must be formally given prior to data deposit (<http://eidc.ceh.ac.uk/deposit/service-agreement-template>).

As part of the agreement, depositors' responsibilities are detailed. These include:

ensuring that in the case of research data with human subjects consents collected are ethically and legally appropriate and sufficient to allow deposit of the resource.

the EIDC must be notified promptly of any copyright, confidentiality, privacy, data protection, defamation or similar issues pertaining to the data.

guaranteeing that nothing in the resource or supporting information contravenes the Data Protection Act 1998 or any other EU/UK law.

Once data is received from the depositor (and prior to it being made publicly available) a number of Resource Acceptance Checks are carried out by data centre staff. 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 EIDC does not normally accept data where there is a disclosure risk and ask that any data containing sensitive information be anonymised or be of a sufficiently broad scale that sensitive location information is not made publicly available.

In the unlikely event that data is made available that contains sensitive information, the EIDC has a withdrawal policy (<http://eidc.ceh.ac.uk/doipolicy>) and a new version of the dataset without the sensitive information would be made available.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

V. 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 EIDC is funded through the NERC National Capability Data Centre Commissioning project. The funding is for an initial period of five years and the EIDC reports annually to NERC on the number and quality of the agreed functions and services. The EIDC currently has a FTE of 12 permanent staff, three of which are project funded.

As part of the process to ensure continued funding of the NERC data centres, a commissioning process was initiated in 2016 that included a stakeholder survey to evaluate the services each of the data centres, including the EIDC, must provide to its designated community for the future. The results of this survey have been used to guide the priorities and

services delivered to users by the EIDC. The outcomes of this process also form the basis for planning future stakeholder engagement activities that will include user surveys and mechanisms to provide feedback on services delivered by the EIDC e.g. web-based feedback forms.

The Centre for Ecology and Hydrology (CEH) is a multidisciplinary centre of excellence for terrestrial, freshwater and related atmospheric science which creates and hosts many long-term, large-scale datasets of national and international importance. NERC provides the long term National Capability (NC) funds to provide the Environmental Informatics Data Centre (EIDC) to support all NERC terrestrial and freshwater science (TFS) in data management, data integration and data dissemination. CEH hosting of EIDC enables experienced data specialists to work closely with CEH scientists and HEI research partners. This has enabled a strong understanding of the needs of researchers across TFS disciplines and experience of co-delivery of data-focussed projects. As a result, EIDC constitutes a highly efficient and effective critical mass of expertise in data management and data science.

EIDC staff are experienced environmental data specialists (data managers, data scientists, web developers and software engineers) with in-depth knowledge of the science they support. EIDC staff continually review and improve the efficiency of technologies and procedures from ingestion to delivery. These staff have expertise in a range of skills including digital preservation, scientific data accession, active data management planning, information architecture, international and regional data standards and web services. The EIDC utilises differing proportions of effort from a range of these staff in the course of delivering its functions and services.

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

The EIDC governance arrangements set the strategic direction, allocating resources, reviewing progress and making decisions. The Head of Environmental Informatics report annually to NERC on the delivery of the agreed services and functions identified in the NERC Data Centre Commissioning Proposal.

Strategy & Finance - The CEH Head of Environmental Informatics (EI) leads development of EIDC strategy which is approved by NERC on a 5 yearly basis and reviewed annually.

Project delivery and quality assurance - Activities are delivered through defined projects with Project Managers reporting to the Head of EI. The EIDC data centre project was examined by BSI assessors in May 2016 as part of CEH's successful accreditation for ISO 9001.

Independent advice - Senior CEH scientists and IT experts are members of an EI Advisory Group which meets quarterly and ensures EI development priorities are informed by science needs. EIDC activities are evaluated by the external National Capability Data Service (NCDS) steering committee which meets annually.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

VI. 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:

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:

Staff that manage the EIDC are employees of UK Research and Innovation (UKRI) - Centre for Ecology & Hydrology (CEH). CEH encourages and supports on-going employee training and relevant accreditation to ensure that all staff have appropriate and up-to-date knowledge and skills. Staff are required to set at least one new training and development objective annually and an appraisal is held bi-annually to ensure training and development requirements are being met.

The EIDC staff proactively engage with a number of initiatives and organisations that provide expertise on a range of relevant topics to ensure that the services provided by the data centre are in line with the current best industry practice. This includes bodies such as the Research Data Alliance (RDA) where EIDC staff are members of several interest and working groups (e.g. Vocabulary services interest group, data type registries WG, Data usage metrics WG, Data citation WG) and metadata working groups (e.g. NERC metadata quality assurance group, DataCite [via the British Library] and AGI GEMINI working group).

The EIDC has provided support and training to the environmental science community in:

Data science skills - through HEI teaching and placements (e.g. teaching on Lancaster University Environmental Informatics MSc; hosting four 3-month placements of MSc Data Science students within EIDC team)

Data management planning, guidance & support (e.g. direct support to HEI staff as part of data management planning and dataset ingestion (also at <http://eidc.ceh.ac.uk/support/dataManagementGuidance>)

Data management delivery (e.g. training courses in principles of data management and database fundamentals delivered to CEH staff)

Promoting upskilling of researchers in novel data analysis approaches by providing tools to explore and work with datasets along with guidance on their use (e.g. drought portal: <https://eip.ceh.ac.uk/droughts>)

Staff at the EIDC are also actively engaged with organisations collecting data citation metrics (Google and Clarivate Analytics) and various publishers of data papers e.g. Nature (Scientific data), The British Ecological Society, Earth System Science Data and Elsevier etc.

EIDC staff regularly interact with staff from other NERC data centres (National Geoscience Data Centre at BGS, Polar Data Centre at BAS, British Oceanographic Data Center at NOC and the Centre for Environmental Data Analysis at STFC) both on an informal basis and also through a formal NERC Data Operations Group (DOG) that coordinates and advises on various aspects of data management policy across the entire research council.

The EIDC is funded to deliver core data centre services as well as a small amount of resource to investigate and implement new technologies for the data centre. The EIDC's priority is to deliver the core services expected of a NERC data centre requiring a division of resources between maintaining the existing infrastructure and developing new capabilities. Input from community experts is captured and managed by an issue tracking system (JIRA) and prioritised by the Head of Environmental informatics and the Environmental Informatics Advisory Group.

In addition, Science Information Strategy is led by the NERC Information Strategy Group (ISG) which comprises Directors of Institutes, Heads of NERC Data Centres and external advisors; they are currently developing the new Science Information Strategy for the next five years. The head of the EIDC is a member of the NERC Information Strategy Group 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 the EIDC is provided by ISG and implemented across the NERC Data Centres through the NERC Data Operations Group. Internal management of the EIDC is through the EIDC operation group.

The ISG recognises that groups with similar aims exist at both the national and international level. For example Jisc (formerly the Joint Information Systems Committee) provides strategic information leadership on behalf of the UK Higher Educational establishments. The ISG liaises with and provides input to these groups where appropriate.

All staff working on EIDC tasks are also involved in other research projects in the organisation, predominantly as data scientists, data managers and informatics liaison officers, which ensures they are up-to-date with the latest technologies, aware of user's developing requirements and aware of change in the research data environment.

Researchers who deposit data with the EIDC are asked to provide feedback to the data centre via a survey which can be accessed by clicking a link provided to them at the end of an ingestion. Any feedback received is collected in the EIDC's issue tracking system (JIRA) where it can be collated into a report for regular review.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

DIGITAL OBJECT MANAGEMENT

VII. 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

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

The EIDC has a data acquisition policy in place that sets out the requirements for depositing data with the data centre (<http://eidc.ceh.ac.uk/policies/acquisition>), which include generation of UK Government GEMINI standard compliant metadata, terms, and conditions.

Metadata and supporting documentation, describing the dataset to facilitate its re-use, must be completed prior to the data being published and receiving a DOI. Processes followed by the Data Centre Operatives and automated systems put in place ensure metadata reaches a sufficient standard prior to publication (e.g. a DOI cannot be issued unless the mandatory DataCite metadata fields are completed in the catalogue). All catalogue records are under Git version control, thus providing an audit trail of alterations to metadata records, and of the user who made the changes. Supporting documents must include information covering relevant areas described in the guidance notes to NERC's data policy, to enable re-use of the data without recourse to the original author, and provide additional provenance data over and above that collected in the catalogue record.

Data and supporting documentation are deposited via the EIDC data uploading tool (available through the catalogue) which automatically stores the data in the EIDC data store and creates checksums for each file, to be used for confirmation of data integrity. Data and supporting documentation are checked by Data Centre Operatives (Resource Acceptance checks) to ensure they are of the correct size, format and provide adequate description of the data to enable future re-use. Trained Data Centre Operatives and Informatics Liaison staff ensure data being deposited is suitable prior to deposit, therefore, once deposited, no further changes are made to the data and verification of the checksums created on deposit attest to this. The EIDC is in the process of implementing an automated tool that will automatically test all the checksums for data held by the EIDC and produce a regular report providing warnings if changes to the data are detected. Supporting documentation may be changed after deposit of the data. Supporting documentation is currently stored in a Plone content management system which provides a full history of any changes that have been made and by whom. Data Centre staff may improve the supporting documentation as part of a continual process of improvement, sometimes resulting from feedback from users or at the request of the depositor themselves.

The EIDC has a publicly available DOI policy (<http://eidc.ceh.ac.uk/policies/doipolicy>) which makes clear that no changes can be made to data once a DOI has been issued. If a dataset is later found to contain errors or the depositor wants to change it in any way a new ingestion must take place with a new metadata record and a new DOI. The dataset being replaced still remains available for users to access, however, direct download will be removed and users would need to email the EIDC directly for access to the original dataset. The new version of the dataset will appear with a version number after the title and links are provided from the superseded dataset to the replacement dataset. A description of why the dataset was superseded is also included in the metadata of the catalogue record.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VIII. 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 deposit requests to the EIDC are fed through our Identification process to determine whether a particular data resource is suitable for ingestion. A single deposit request may be made for one or many datasets, however assessment of suitability for ingestion is made on a dataset by dataset basis, with an individual Identification Checklist being completed for each dataset. Once completed, the checklists are stored in a Quality Log, regardless of the outcome. The Identification checklist is based on the NERC Data Value Checklist

(<https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/>) and identifies a title, file size and current format and whether we already hold related datasets. The checklist asks questions such as: Are the data adding to, replacing, or otherwise affecting existing EIDC data holdings; are the data NERC funded; do the data cover likely areas of future scientific interest; is the EIDC the most appropriate data centre for deposit; are the data unique; are the data repeatable and are the data reusable? A Data Centre Operative will complete the checklist and this will determine whether the data are suitable for deposit at the EIDC. There are three possible outcomes: the data are not considered suitable for deposit; the data are not considered suitable for deposit with the EIDC but we can recommend an alternative repository for the dataset (another NERC Data Centre or more specialist subject specific repository, for example); the data are suitable to progress through our deposit process. The latter does not guarantee that the EIDC will take the data, merely that it is the

most suitable data centre to consider taking them and the process of deposit can commence.

The EIDC has a publicly available acquisition policy (<http://eidc.ceh.ac.uk/policies/acquisition>) to guide potential depositors on our remit.

Once a data resource has been identified as suitable for deposit, the EIDC will work with the depositor to draft a Service Agreement (<http://eidc.ceh.ac.uk/deposit/service-agreement-template>) which documents details about the dataset, for example, file format, licencing, embargo details, details to mint a DOI and what will be provided in terms of supporting documentation, etc. Depositors receive guidance from Data Centre operatives and information is available on our website as to what is expected from them. Accepted file formats and naming conventions are described here:

<http://eidc.ceh.ac.uk/help/depositing/preparingForDeposit> and

<http://eidc.ceh.ac.uk/help/depositing/guidance-suitable-file-formats> guidance on providing supporting documentation is provided here: <http://eidc.ceh.ac.uk/deposit/supportingDocumentation> and guidance on the metadata we require is provided here: <http://eidc.ceh.ac.uk/deposit/metadata-guidance> .

The EIDC prefers to ingest data in non-proprietary formats. However, our list of preferred formats is updated regularly and we will work with depositors offering data in a non-preferred format to reach a compromise if possible. Non-proprietary formats are also preferred for the supporting documents which are made available alongside the resource to facilitate its re-use.

Once data and documentation have been deposited with the EIDC via the agreed method (most commonly via our web-based data uploader facility) they are checked against our internal standards and the details agreed with the depositor in the Service Agreement. A Data Centre Operative will complete a Resource Acceptance Checklist for each deposit, which is then stored securely for future reference. The Resource Acceptance checklist checks file names, format and sizes are as expected, that the files open correctly and contain appropriate content. We also check the data doesn't contain information contravening the Data Protection Act. The supporting documentation is checked against the data itself to ensure it is sufficient to enable re-use of the data (e.g. files, methods, column names, codes and units are described). We also do a basic spelling and grammar check and ensure any references are properly cited at the end of the document. If the Resource Acceptance Checks are not passed, we contact the depositor explaining how the data/documentation have failed the checks and ask them to resubmit the data/documents after corrections have been made. The files which have been found to contain errors are deleted.

As well as data and supporting documentation, the depositor must provide sufficient information to complete a discovery metadata record for the data. Depositors can create a user account with the EIDC and log in to the metadata catalogue to do this themselves, or Data Centre Operatives can complete the record on their behalf if they provide the necessary details via email. The metadata record is checked automatically and warnings are displayed on-screen if mandatory elements are missing. The metadata quality checking is based on the INSPIRE/UK government UK GEMINI metadata standard and includes extra mandatory fields required by the EIDC. The EIDC does not accept discipline specific metadata profiles given the need to accommodate records for many different types of terrestrial ecological and hydrological data within a single data catalogue. Prior to publication, a further manual assessment of the quality of the

metadata record is made by a Data Centre Operative. Discovery metadata records are not published to the public until all the warnings/quality issues have been satisfactorily resolved.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

IX. Documented storage procedures

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

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The EIDC 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).

Ingestion of data is managed via a series of documented processes. Processes are under change control, meaning any alterations must be approved by the data centre management team and quality assured prior to going live. Secure access for EIDC staff to processes is provided via a password protected admin section of the EIDC website. Processes and their associated guidance are under continual process improvement. Feedback from depositors and data centre staff and analysis of non-conformance reports precipitate changes to the processes where improvements can be made.

The repository follows the best practice guidance as described in the preservation policy (<http://eidc.ceh.ac.uk/policies/digitalPreservation>). This includes the ISO14721 (OAIS Model) for storage as well as the other preservation functions. Full details of the EIDC's archival storage function can be found in the digital preservation policy. Briefly, files are backed up on tape incrementally daily with a full back-up every month. As part of the EIDC's disaster recovery plan, tapes are stored off-site in a secure facility. As such, if any corruption of deposited data files is discovered, backed-up copies can be retrieved. Back-ups of the system are managed by local IT support (CEH Computer Support), who carry out test retrieval of data to ensure back up tapes are readable on an annual basis.

Once the original data (Submission Information Package (SIP)) has been uploaded via the EIDC's web-based data uploader it is stored securely on the EIDC Storage Area Network, behind a firewall. Fixity values (checksums) are created for each data resource and validated on a regular basis. Depositors are required to take responsibility for uploading the data they are depositing – data centre staff are not permitted to upload it on their behalf, in order to ensure the files received are the correct files. Once data and supporting documentation have been received it is checked against what was previously agreed in the Service Agreement (file names, format, size etc). Supporting metadata documents are stored on the EIDC website, which as a content management system also provides a history of uploaded documents, enabling roll back if required.

Authorisation of access to the Storage Area Network which holds the data is limited via permissions to the minimum viable number of staff. These permissions can only be updated by a member of the CEH Computer Support team. Access to the network is also password protected. The building housing the repository is equipped with a security-protected card access system.

The EIDC 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. If a change were identified a report would be automatically emailed to the EIDC's issue tracking system to alert staff of the problem.

The CEH Risk Register documents significant corporate risks using a scoring system, and outlines appropriate mitigation scenarios. The risk register is monitored and updated on a regular basis.

Disaster recovery procedures are in place that include data recovery provisions involving restoring data from tapes and controlled restoration of EIDC services e.g. catalogue, website etc.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

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

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 EIDC's role as the national terrestrial and freshwater 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. However, the majority of the data held by the EIDC has been assigned a Digital Object Identifier (DOI), therefore we guarantee to hold it in perpetuity. The responsibility for the data is transferred from the depositor to the EIDC upon deposit. Should CEH cease to exist as an organisation, UKRI would have long-term responsibility for all the EIDC data holdings.

The EIDC has a digital preservation policy (<http://eidc.ceh.ac.uk/policies/digitalPreservation>) which documents the model under which the data centre operates, including preservation planning, roles and responsibilities, funding and the legal and regulatory frameworks the data centre is bound by.

The 'preservation level' for each data resource is the same. For all datasets, a Service Agreement (<http://eidc.ceh.ac.uk/deposit/service-agreement-template>) is drawn up between the data centre and the depositor, describing each party's responsibilities. The Service Agreement documents 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 (e.g. embargo period, creation of web map services) and details on licencing and IPR.

The data and supporting documentation are not accepted by the data centre unless the Service Agreement has been formally agreed by both parties. Once a Service Agreement has been agreed, a date for submission of the data is set. Upon receipt, the EIDC assumes responsibility for the data resource. The Service Agreement clearly states: 'The EIDC will store the data and, where necessary, duplicate and transform copies' and 'We will make the data publicly available, subject to any conditions specified in this Service Agreement.' Data and supporting documentation deposited with the EIDC have to meet specified Resource Acceptance Checks prior to being moved to the area of the data store reserved for long-term secure storage. These checks ensure that the data and documentation deposited are as agreed in the Service Agreement. Guidance is available to depositors on the submission standards required by the data centre (<http://eidc.ceh.ac.uk/deposit>). A record of the checks performed is held by the data centre and a quality log for each deposit holds records of quality assurance checklists completed after each stage of the deposit process. A final, overall quality assurance check of all work undertaken as part of the deposit of data to the data centre is made, to ensure that all work has been completed satisfactorily and relevant actions have been taken, the results of which are stored in the Quality Log. The EIDC manager, or other nominated individual, will make a daily check of all postdated tasks, such as lifting of embargos, to ensure they are completed in a timely manner. Management of postdated tasks and all other tasks associated with the deposit of data to the data centre is managed using a Task Management System (JIRA).

The EIDC has a well-established data management planning procedure in place for both internal and NERC grant data that takes into account the whole data lifecycle (<https://nerc.ukri.org/research/sites/data/dmp/>).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XI. 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:

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:

It is not feasible for the data centre to check the scientific quality of the data it receives, given the diverse range of research areas it covers and the variety of data resource types it ingests. NERC expects the scientific quality of the data generated by its funded grants and programmes to be assured by the scientific staff delivering such programmes. However, the data centre does conduct several 'sanity' checks upon receipt of data. Such checks include: data files being in the correct format; data files open and are readable in an appropriate application; the content of the file(s) is as described in the metadata; data types are appropriate and consistent; max/min values are not exceeded (where relevant); nulls/missing values are recorded appropriately and consistently; the data contains no information of a personal nature (e.g. names & email addresses). All data centre staff have backgrounds in environmental data science, and will use their experience to judge whether the content of a data file is sensible, though as stated previously, they are unable to make an assessment of scientific quality of the data.

The EIDC requires that all deposits are accompanied by comprehensive discovery metadata. Discovery metadata are assessed against a set of quality rules incorporating the UK GEMINI metadata standard (<https://www.agi.org.uk/agi-groups/standards-committee/uk-gemini>) and the NERC metadata quality guidelines. It is also subject to a regular appraisal conducted by external reviewers organised by the NERC Data Operations Group. The EIDC data catalogue has in-built validation, which will verify the record content against the relevant metadata standard e.g. UK GEMINI for European spatial datasets. Results of the validation are available when a record is saved, providing details of outright errors requiring correction, and warnings where a problem is suspected.

Discovery metadata records are stored as json documents in a git repository; all changes are therefore under revision control.

Discovery metadata records are publicly available via a searchable catalogue at <https://catalogue.ceh.ac.uk/eidc/documents> . Records are available in a number of formats: json, GEMINI xml, RDF (TTL) and Datacite xml. They are also harvested and made available through a number of other systems including the NERC Data Catalogue service (<http://data-search.nerc.ac.uk/>), Find Open Data (<https://data.gov.uk/>) and Google Dataset Search (<https://toolbox.google.com/datasetsearch/>), with relevant records also being harvested by the EU INSPIRE Geoportal (<http://inspire-geoportal.ec.europa.eu/>).

Discovery metadata records all contain details of how to cite the resource, but may also contain links to other related data resources and/or documents e.g. academic papers which have used the resource, grey literature related to the resource, project websites, etc., though these are not mandated, as the data centre is primarily concerned with the data resource itself.

In addition to discovery metadata, the EIDC also requires that data is accompanied by descriptive metadata to help interpret and reuse the data. In short, this supporting metadata should be sufficiently detailed as to enable re-use of the resource by a member of the community without further recourse to the authors of the data or the data centre. Such documentation must include an account of the data structure with a description of all variables including data types and units of measurement. It is also likely to include details of: the experimental design/sampling regime; data collection methods; data transformation methods; fieldwork/laboratory instrumentation used; analytical methods and any quality control measures employed. Guidance on producing such metadata is available on the EIDC website at <http://eidc.ceh.ac.uk/deposit/supportingDocumentation>. The descriptive metadata is held in a version-controlled repository and is available for download separately from the data resource itself in order to permit users to make an informed decision regarding the data prior to agreeing any licence and/or initiating an order/download of the data. The EIDC reserves the right to amend or enhance the supporting metadata documents at any time after the deposit is complete.

Data users have the option to comment on data and/or metadata supplied by the EIDC via completion of our contact form accessed via the data catalogue. Contact in this manner has led to several useful amendments or corrections to resources we hold. The data centre used to operate a 'Rate this dataset' function, allowing users to provide a star-rating for each resource. However, this facility was discontinued after a number of years as it was never used.

The EIDC is working closely with DataCite, Clarivate and the RDA to enable the automated collection of citation metrics for the datasets it holds. A citations section was recently added to the catalogue (see example: <https://doi.org/10.5285/0995e94d-6d42-40c1-8ed4-5090d82471e1>), however, citations currently need to be added manually by data centre staff. As part of the support data centre and Informatics Liaison staff give to CEH and other NERC funded researchers, depositors are encouraged to consider submitting data papers describing the datasets they have generated in order that they may be peer reviewed and, if published, receive additional citations.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XII. 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:

All ingestion of data to, and provision of data from, the EIDC is undertaken according to defined, quality controlled processes. All processes are documented in a secure admin area of the EIDC website, and are subject to change control, meaning any alteration to business processes must be approved by data centre management and quality assured before becoming part of the working practices. All Data Centre Operatives conducting work ingesting resources to the EIDC adhere closely to the documented processes. Deviations from the processes are identified during Quality Assurance, and

are raised as non-conformances to be addressed by the data centre management team. Communications with depositors of data are made using the official EIDC email account, using standard templates which are amended with information unique to the relevant ingestion. All quality assurance of completed tasks is undertaken via completion of the relevant checklist, which is stored securely in the Quality Log folder of the admin area on the EIDC website.

Ingestion of resources follows a clearly defined process from start to finish, details of which are available to data depositors on the EIDC website (<http://eidc.ceh.ac.uk/deposit/deposit-process>).

The process begins with a request to deposit data from an individual, henceforth known as the depositor. This triggers a Data Centre Operative to complete the Identification process in order to assess suitability of resources for ingestion to the EIDC; the outcome of which is communicated to the depositor via email. Identification is made via completion of an identification checklist per resource requested for deposit, which is based on the NERC Value checklist (<https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/>), and queries issues such as long-term value, reusability, uniqueness, etc. If a resource is identified as not suitable for ingestion by the EIDC, the individual requesting deposit will be notified via email and given reasons for the decision. If possible, another suitable data repository is recommended to the depositor. If a resource is identified as suitable for ingestion by the EIDC, an ingestion 'job' for the deposit will be created within the EIDC's task management system. The 'job' will be picked up by one of the EIDC's ingestion managers, who will follow the Ingestion Management process to co-ordinate the activities involved in ingestion of all suitable resources to the EIDC.

The ingestion manager will quality assure that identification has been completed correctly, and will then assign a Preparation task for each resource identified to a Data Centre Operative, who will follow the Preparation process. As part of the Preparation process, a Service Agreement is drafted, detailing what data and documentation the EIDC can expect to receive from the depositor, and what services the depositor can expect to receive from the EIDC (<http://eidc.ceh.ac.uk/deposit/service-agreement-template>). The Service Agreement specifies what files are to form the deposit, what supporting documents will be supplied, the format and size of files, details of licensing, embargo and any copyright statement required and the details of individuals to be identified as authors of the resource for the purposes of registering a DOI for the dataset. Once a Service Agreement has been drawn up, a copy is emailed to the depositor for their agreement, which is indicated by return of email. Additional authors identified on the Service Agreement are emailed to inform them that they have been listed as an author of the data resource to be published. A record is subsequently created in the EIDC metadata catalogue containing appropriate details from the Service Agreement, such as title and authors, although personal author email addresses are not displayed. Creation of the catalogue record also generates a unique identifier for a resource, which is used internally within the data centre to identify the resource and tag issues relating to work undertaken for the ingestion of the resource.

Once complete, the work undertaken as part of the Preparation task will be quality assured by the ingestion manager. A Data Transfer 'task' for each resource to be ingested is then set up. The Data Centre Operative completing the Data Transfer task will agree a date for transfer of the data resource and supporting documents to the data centre. Data and supporting documentation are typically deposited via the file uploader on the EIDC catalogue. The file uploader is only accessible to registered users of the catalogue who have been assigned the correct permissions by data centre staff.

Larger files may require an alternative method of transfer, such as transfer via FTP site or physical storage device.

Once transferred, data resources and supporting documents are subjected to Resource Acceptance Checks, which assess whether the transferred files are as expected, whether the files open correctly and contain appropriate data values, if all variables and codes used within the data are unambiguously described in supporting documents accompanying the data resource, etc. If the checks are passed, the Data Centre Operative will move the files to the correct location for their secure storage. The depositor is sent a receipt email, to notify them that the files have been received and are in agreement with the files listed on the Service Agreement. A copy of the checksums that the EIDC has generated for the files is also sent to the depositor, so that they can verify that the files have not corrupted during transfer. If the Resource Acceptance checks are failed, the depositor will be informed and asked to submit corrected files.

During the Data transfer task, the Data Centre Operative will also ensure that the discovery metadata catalogue record is completed to the relevant standard prior to publishing the record. The ingestion manager will quality assure the work that has been undertaken, before minting a DOI for the resource and notifying the depositor. If an embargo has been requested, the ingestion manager will update the metadata record with the embargo details, and create a post-dated task to lift the Embargo, to be completed once the embargo has expired. Otherwise, they will create a task to Enable Online Access to the resource, and assign it to a Data Centre Operative.

The Data Centre Operative will follow the 'Enable Online Access' process, to include testing of the delivery mechanism. At the end of this task, the data and supporting documentation will be made publicly available to users under whichever licensing conditions were agreed as part of the Service Agreement. The Ingestion Manager will Quality Assure the work, prior to creating a 'Tweet task', which will be assigned to a Data Centre Operative. This staff member will complete the task by 'tweeting' that the data resource is now available on the EIDC Twitter account (@CEH_EIDC), and assign the task back to the Ingestion Manager for Quality Assurance.

Further additional tasks also form part of the ingestion workflow, depending on the requirements of the depositor agreed when the Service Agreement was created, those tasks are performed less frequently, and do not form part of the standard process of ingestion. These can include the requirement to create an Aggregate or Series catalogue record in the metadata catalogue, to act as an umbrella record for related data resource records, or to supersede a dataset as a result either additional data being added to an existing dataset or discovery of errors within an ingested dataset.

Following completion of all tasks forming part of the ingestion of the resources to the EIDC, with the exception of the post-dated tasks created for data under embargo, the management of the ingestion by the ingestion manager is Quality Assured by a member of the management team who was not involved in undertaking any of the work involved in the ingestion.

Reporting of data centre outputs to stakeholders is made from filtering the task management system (JIRA) to determine data resources published on a monthly and annual basis, which can be further broken down by scientific discipline.

Where improvements or errors within the ingestion workflow are identified, a change request issue type is created within

the data centre's issue management system, detailing the improvement or problem, and the work needed to effect the required change. This is reviewed by a member of the data centre management team, who will approve the work and assign it to a member of staff to complete, or reject the change, giving full explanatory reasons behind the decision. All changes are quality assured prior to becoming working practice.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIII. 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:

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 EIDC provides access to the data resources it holds via a bespoke discovery metadata catalogue (<https://catalogue.ceh.ac.uk/eidc/documents>). Data resources are described to the INSPIRE/UK government UK GEMINI metadata standard (https://guidance.data.gov.uk/publish_and_manage_data/harvest_or_add_data/harvest_data/gemini/#iso-19139), with the exception of a small number of data resources with no geographic element e.g. purely lab-based studies. Catalogue records are also harvested by metadata aggregator sites such as the NERC Data catalogue Service, Data.Gov UK and EU JRC.

The EIDC is listed in the Registry of Research Data Repositories (<http://doi.org/10.17616/R3XG68>), further facilitating exposure for the data resources held by the EIDC.

The EIDC catalogue allows users to search for data resources of interest via keywords or spatial extent and to further filter their results based on the facets topic, record type, availability and licence. An EIDC catalogue record displays information about the data resource (description, accessibility, temporal and spatial extent etc.) and also displays a 'You must cite' panel which details the recommended citation for the data (e.g. <https://doi.org/10.5285/5321bc6e-be35-4ed3-9b56-25598d61ac8f>).

The EIDC is an issuing agent for DataCite DOIs (<https://www.datacite.org/>). The data centre has a documented process for assigning DOIs and all data resources accepted for deposit will usually be assigned one. A DOI is a unique identifier for a data resource, which can be dereferenced in a web-browser to direct the user to a 'landing page' containing metadata about the data resource. The EIDC uses the catalogue record for the dataset as the 'landing page' for the DOIs it issues. In order for the EIDC to assign a DOI to a data resource it must be fully ingested into the data centre to ensure it is of the required quality, and has all of the necessary supporting information to enable re-use.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

XIV. 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

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Discovery metadata complying with INSPIRE 19115/19139, which in the UK also complies with the UK GEMINI v2.1 schema is required when data is deposited to the EIDC. 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. data.gov.uk and the NERC data catalogue service. 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.

The EIDC provides guidance on the file formats accepted by the data centre (<http://eidc.ceh.ac.uk/help/depositing/preparingForDeposit> and <http://eidc.ceh.ac.uk/help/depositing/guidance-suitable-file-formats>) to encourage deposit of data in formats that are at less risk from technology/software obsolescence, or require costly proprietary software to open. The preferred formats are generally non-proprietary, or open, industry-standard formats and can be used by anyone wishing to access the data. For example, a depositor holding their data in an MS Excel spreadsheet would be asked to deposit the data as a .csv file. Other accepted file formats are .txt, .dat, .gdb, .tiff and .shp.

An annual review is undertaken to take account for the possible evolution of formats, however, holding data in mostly non-proprietary format minimises the risk of a migration to another format being required. In the event of a planned migration we would use the formal configuration management and change control process that is applied to anything requiring a change within the data centre (e.g. supporting documentation, metadata, process) First, a change request would be generated in the tracking system (JIRA) and the files requiring migration would be identified. Second, the migration would be performed using the tracking system (JIRA) to track progress; finally, the migrated files would be checked to ensure usability. After the migration was complete, the change request would be quality assured by another member of staff to ensure the migration had been carried out satisfactorily.

In order to ensure continued understanding of the data, detailed supporting documentation (in addition to discovery metadata) is captured at the data ingestion stage. The EIDC recognises the importance of providing as much contextual information as possible for its data holdings in order to facilitate its re-use. As much supporting information as possible is captured alongside the raw data, and deposited with the data centre at the same time as the data resource itself. Provision of this additional supporting documentation, available to download separately from the data itself, enables users to make an informed assessment of whether individual datasets are fit for their specific purpose and to make appropriate use of it, without recourse to the data authors. Supporting documentation is available to download from the EIDC catalogue record for the dataset.

The EIDC provides guidance to depositors as to the depth of information required in the supporting documentation

(<http://eidc.ceh.ac.uk/deposit/supportingDocumentation>). Depositors are required to list the file names and formats of the supporting information they will be supplying and the areas the document cover, specifically, experimental design/sampling regime, collection methods, fieldwork and laboratory instrumentation, calibration steps and values, analytical methods, nature and units of recorded values, quality control and details of data structure, in the Service Agreement prior to deposit.

Upon receipt of the data and supporting documentation, data centre staff check the contents of the supporting documentation against the data deposit as part of the Resource Acceptance Checklist. Data centre staff are required to check that the supporting documentation is sufficient to allow re-use of the data, specifically checking that the files submitted, methods, column names, codes and units are all fully described. Staff also do a basic grammar and spelling check and ensure any references mentioned in the text are listed in full at the end of the document.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. 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

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:

Online storage for the data centre repository is provided by a Storage Area Network (SAN) administered by a dedicated in-house IT infrastructure team. Data is backed up to Enterprise Tape Libraries continually daily. The backups are managed using IBM Tivoli Storage Manager (TSM) - tapes are stored off-site in a secure facility.

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 CEH computer support team maintains a software inventory, including documentation covering the local IT infrastructure that is maintained internally.

The EIDC has a dedicated development team whose role is to ensure that the infrastructure, software applications and web frameworks are kept up-to-date and maintain the functionality required to deliver a NERC Data Centre. This also involves identifying new technologies that would increase efficiencies or add additional capabilities as required e.g. automating ingestion of data. This feeds into the IT component of the Data Centre Strategy mentioned in R6.

The dedicated software team manage the development using agile methods and these developments are documented using a Wiki and managed with our issue tracking systems (JIRA). This way we can keep abreast of our systems.

The data centre metadata catalogue has been developed in-house and the code is made publicly available on GitHub under an open source GNU licence (<https://github.com/NERC-CEH/catalogue>). It is hosted on a VMware virtual infrastructure, maintained using software configuration management (Puppet) and documented using Terraform to ensure it is easily replaceable and reproducible in the event of a disaster. Formal maintenance agreements are in place for VMware. The EIDC is connected to the UK Joint Academic Network (Janet) over 1Gbps link. In the near future, a dual-redundant link will be established to maximise service availability.

Discovery metadata conforms to the UK GEMINI schema (<https://www.agi.org.uk/agi-groups/standards-committee/uk-gemini>) and is thus also compliant with INSPIRE metadata requirements.

Delivery of small datasets is performed via email; larger data volumes are streamed over http. The EIDC also provides access to some data via OGC compliant web services and custom APIs.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XVI. 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

Reviewer 2

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

Response:

The EIDC repository has a comprehensive set of procedures in place to ensure rapid recovery and return to normal operations in the event of a system failure (see also infrastructure documentation in section R15).

All data is stored on Storage Area Networks at Lancaster and are available as Microsoft Windows file shares in an Active Directory managed by the Centre for Ecology & Hydrology (CEH). Access to the SAN and file shares is controlled with Check Point Firewalls and external access is only available via a Secured VPN with multi-factor authentication using digital tokens.

Access to the data is authorised using Active Directory file permissions that allow only privileged users to access the data (read-only or read-write as appropriate). In addition, Carbon Black Protection is used to monitor the SAN and all endpoints; it blocks any malicious activity that could potentially corrupt or damage data.

All data is continually backed-up using IBM Tivoli Storage Manager. The data are part of a retention and recovery schedule that allows a rolling twelve months' worth of file retrieval. A copy of the archive is stored on tape and securely kept off site, in line with the EIDC's disaster recovery policies.

The EIDC utilises the expertise and skills of CEH's Head of Information Security. They are responsible for advising on

implementation and maintenance of the security of CEH and EIDC infrastructure. Functionally they provide technical support for all aspects of CEH/EIDC cybersecurity. This includes configuration of the firewalls and provision of data access for external users via the secure VPN.

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). The system also ensures the safety of the data in the event of a more serious incident if, for example, the buildings housing the data centre and/or major IT infrastructure were to be rendered inoperable.

CEH has Cyber Essentials accreditation (<https://www.cyberaware.gov.uk/cyberessentials>) and is also certified to ISO:9001.

EIDC process is in place to ensure data held under embargo is not available to the public until the agreed end of the embargo period. Data held under embargo are stored in the EIDC secure data storage area where they remain inaccessible to the public. A post-dated task is created in the task management system (JIRA) which alerts the EIDC when an embargo is due to expire. After this date the task is assigned to a Data Centre Operative to make the data resource publicly available and accessible through the catalogue. Data Centre processes ensure that the operative charged with the task of lifting the embargo double-checks the date of expiry to ensure data are not made public prematurely.

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 core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: