



## **Implementation of the Data Seal of Approval**

The Data Seal of Approval board hereby confirms that the Trusted Digital repository Edinburgh DataShare complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.

The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on October 22, 2015.

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

Yours sincerely,

The Data Seal of Approval Board

## Assessment Information

|   |   |
|---|---|
| Guidelines Version:   | 2014-2017   July 19, 2013   |
| Guidelines Information Booklet:   | <a href="#">DSA-booklet_2014-2017.pdf</a>   |
| All Guidelines Documentation:   | <a href="#">Documentation</a>   |
| Repository:   | Edinburgh DataShare   |
| Seal Acquiry Date:  | Oct. 22, 2015   |
| For the latest version of the awarded DSA for this repository please visit our website: | <a href="http://assessment.datasealofapproval.org/seals/">http://assessment.datasealofapproval.org/seals/</a> |
| Previously Acquired Seals:  | None  |
| This repository is owned by:  | <ul style="list-style-type: none"><li>• <b>EDINA &amp; Data Library, University of Edinburgh</b></li></ul>    |
|   | Scotland  |
|   | T +44 131 651 1431  |
|   | E <a href="mailto:datilib@ed.ac.uk">datilib@ed.ac.uk</a>  |
|   | W <a href="http://www.ed.ac.uk/is/data-library">http://www.ed.ac.uk/is/data-library</a>                       |

# Assessment

## 0. Repository Context

### Applicant Entry

#### *Self-assessment statement:*

The Data Library at the University of Edinburgh assists staff and students in the discovery, access, use and management of datasets for research and teaching and has been in operation since 1983. EDINA provides online services and resources for UK Higher and Further Education. Together “EDINA and Data Library” is a division of Information Services at the University of Edinburgh.

The Data Library manages Edinburgh DataShare, an online multi-disciplinary digital repository of research datasets produced by researchers at the University of Edinburgh. Edinburgh DataShare is designed to be scalable and highly resilient allowing University of Edinburgh researchers to publish, share, describe, embargo, and licence their data assets for discovery and use by others via the Internet. The repository includes a standards-compliant metadata schema compatible with repository harvesting protocols, a user interface for deposit and administration, search and browse facilities, item-level and file-level usage statistics, time-stamped submissions and persistent identifiers.

The service, established in 2008 and now part of the University’s Research Data Management (RDM) Programme, operates according to a service level definition and a set of policies, which are available from the home page of the repository. These include content policy, submission policy, preservation policy, copyright policies for data and metadata, and a depositor agreement. A public-facing wiki provides additional detailed service information and working documents including developer pages, functional requirements, metadata specifications, and administrative procedures.

We manage access, preservation and re-use of data assets for the researcher as part of the data lifecycle. We do not outsource any aspects of the repository service.

Links to supporting documentation:

- Data Library services: <http://www.ed.ac.uk/is/data-library>
  
- Service background: <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-background>
  
- Edinburgh DataShare repository: <http://datashare.is.ed.ac.uk/>

- Edinburgh DataShare policies:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies>

[Last accessed 1 June 2015]

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

# **1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.**

*Minimum Required Statement of Compliance:*

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

## **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Repository staff guide and assist University of Edinburgh researchers (data producers) from all disciplines with the preparation, description, deposit, embargo, and licensing of their research data assets for safekeeping, online discovery and re-use.

Before depositing data in Edinburgh DataShare researchers are prompted to check and organise their materials (data files, accompanying documentation, metadata) and to consider permissions, rights, disclosure, embargo periods, and licensing. A "Checklist for deposit" is available online to provide initial guidance on data preparation. Data producers may use the checklist to decide on file formats, craft appropriate documentation and ancillary files (README text file containing an outline of the research, methodology, variables, data structure, code book, questionnaires, protocols, etc.), collate information about their data (metadata), and decide whether to select the recommended open licence (CC-BY 4.0) to publish their research data in the repository.

Upon receipt of a new submission, repository staff check that no personal or sensitive data are included, that contextual documentation and metadata provided are sufficient for re-use. Staff also conduct format checks, spot-check for file completeness and readability when possible. Repository staff liaise with data producers to resolve any missing information or inconsistencies. Depositors are encouraged to deposit documentation in an accessible format such as .txt or .pdf, and to provide as much descriptive metadata as possible. Before the submission is accepted into the repository the depositor must amend identified errors in the documentation and address any concerns regarding confidentiality, rights or permissions.

As stated in Edinburgh DataShare's "Submission policy" the validity and authenticity of the submitted data and ancillary files are entirely the responsibility of the data creators and depositor. If the repository receives proof of copyright violation, the relevant item will be removed immediately. When depositors submit their data they agree to the terms of the "Depositor Agreement." As part of the University of Edinburgh, researchers (data producers) must adhere to the "Code of Practice for Research" of the UK Research Integrity Office which covers collection and retention of research data as well as data protection, intellectual property, copyright, licensing, publishing, misconduct, auditing, and other ethical concerns.

**Data Seal of Approval Board**

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

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

Links to supporting documentation:

- Edinburgh DataShare “Checklist for deposit”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/checklist>
  
- Edinburgh DataShare “Submission policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/submission>
  
- Edinburgh DataShare “Depositor Agreement”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/depositor-agreement>

**Data Seal of Approval Board**

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

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

- The University of Edinburgh “Code of Practice for Research”:  
[http://www.research-innovation.ed.ac.uk/Portals/0/Documents/Code-of-Practice-for-Research\\_October2011.pdf](http://www.research-innovation.ed.ac.uk/Portals/0/Documents/Code-of-Practice-for-Research_October2011.pdf)

- UK Research Integrity Office “Code of Practice for Research”:  
<http://www.ukrio.org/publications/code-of-practice-for-research/>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

**Data Seal of Approval Board**

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

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



*Comments:*

## 2. The data producer provides the data in formats recommended by the data repository.

### *Minimum Required Statement of Compliance:*

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

## Applicant Entry

### *Statement of Compliance:*

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

### *Self-assessment statement:*

Edinburgh DataShare is a discipline-agnostic data repository that defines data formats as “recommended”, “selected” and “unknown.” The distinctions and distinguishing criteria are explained in the “Recommended File Formats” document available online.

The data producer can deposit recommended standard preservation formats (“recommended”), and proprietary or discipline-specific formats (“selected” formats have been deposited before and the system recognises them, ‘unknown’ formats are those that are deposited for the first time and are not recognised by the system). “Unknown” formats are evaluated by repository staff and incorporated in the “File Format Registry” once the specifications of the formats are identified and can become “selected”, or in some cases “recommended.”

Edinburgh DataShare provides a “Checklist for deposit” for the data producer which includes recommended file formats to assist in the data preparation and deposit process. The repository also offers tailored online and face-to-face support for data creators. The repository also recommends that data creators revise discipline-specific documentation on standard preservation formats in their respective fields, and to normalise their data from proprietary or discipline-specific to standard preservation formats.

Links to supporting documentation:

- Edinburgh DataShare “Recommended File Formats”: <http://edin.ac/1HbsADs>

- Edinburgh DataShare “File Format Registry”:  
<https://www.wiki.ed.ac.uk/display/datashare/File+format+registry>

- Edinburgh DataShare “Checklist for deposit”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/checklist>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

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

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

### **3. The data producer provides the data together with the metadata requested by the data repository.**

*Minimum Required Statement of Compliance:*

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

#### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Edinburgh DataShare provides a simple submission workflow for depositors to describe and upload their data. Depositors are required to add a minimum set of metadata necessary for e.g. citation. The required metadata fields are marked with a red asterisk (\*). Additional metadata are optional, and are collated through the expandable metadata boxes marked with the plus sign (+). Through the “Depositor’s User Guide” and via tailored support, repository staff encourage and assist depositors to complete optional metadata fields to improve the discoverability and re-use of their research data.

The deposit process is mediated by repository staff and where there is insufficient or erroneous metadata or documentation depositors are contacted and requested to make the relevant amendments. Datasets (defined as data files with accompanying documentation) are ‘held’ for approval by repository administration until the submission is deemed satisfactory. All data files should be accompanied by comprehensive machine-readable contextual documentation such as codebooks, readme files, technical notes, questionnaires, reports, and errata in open and accessible formats.

Edinburgh DataShare operates under a DataCite-compliant DCMI metadata schema which has been found to be robust for a wide range of disciplines. Following a series of test pilots including Informatics, Clinical Psychology, Linguistics and English Language, the University RDM Steering Group declared the repository, with its current metadata schema, is fit for purpose across the university.

Metadata about items in the repository are freely available online and may be reused without prior permission provided that a unique identifier or link to the original metadata record is given, as stated in the repository’s data and metadata policy.

Links to supporting documentation:

- Edinburgh DataShare “Depositor’s User Guide”: <http://edin.ac/1J8VCFX>
  
- Edinburgh DataShare “Data and metadata policy”: <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/data-met>
  
- Edinburgh DataShare “Metadata schema”: <https://www.wiki.ed.ac.uk/display/datashare/metadata>

**Data Seal of Approval Board**

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

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

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

#### **4. The data repository has an explicit mission in the area of digital archiving and promulgates it.**

*Minimum Required Statement of Compliance:*

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

#### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Edinburgh DataShare is hosted by the Data Library. The Data Library assists staff and students in the discovery, access, use and management of datasets for research and teaching. Together with EDINA they are a division of Information Services.

Edinburgh DataShare is a key component in the Research Data Roadmap developed to implement the University Research Data Management Policy. It offers Edinburgh University researchers a long-term preservation service for their research data output and can help researchers comply with funder requirements and with the University's RDM Policy.

Edinburgh DataShare's mission is to anticipate and serve the data sharing, publication and preservation needs of researchers at the University of Edinburgh within an open access environment as detailed on the "Trustworthiness" web page. This complies with the University's mission which is the 'creation, dissemination and curation of knowledge'. The "Preservation policy," "Service level definition" and "Depositor Agreement" demonstrate our commitment to digital archiving. This is promulgated to data producers throughout the university through awareness raising sessions and through, for example the "Benefits of Deposit" page and DataShare promotional leaflet.

As the university's research data repository service, Edinburgh DataShare is available 24x7. Target availability is 99% measured over a year, excluding planned maintenance periods. Core hours of support are 9AM to 5PM during University working days. There is no regular scheduled maintenance period, with all routine maintenance able to be performed without service loss. Any reduction in performance or resilience during maintenance periods will be managed to reduce affect.

Items deposited in Edinburgh DataShare adhere to a DCMI metadata schema and are given a citation, re-use licence and persistent identifier (DOI). The repository complies with OAI-PMH protocol to enable the indexing and harvesting of the metadata and the online dissemination of data records. The repository is listed in the Directory of Open Access Repositories (DOAR) and in the Registry of Research Data Repositories (re3data). Data records are also harvested by Thomson-Reuters Data Citation Index.

**Data Seal of Approval Board**

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

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

Access and download of deposited items are logged through DSpace's statistical tools as well as Google Analytics. Usage statistics are COUNTER-compliant and monitored through Institutional Repository Usage Statistics UK (IRUS-UK), a national aggregation service. Usage statistics are also viewable by depositors and the public at the community, collection, data item and file level within the repository.

Data Library staff are involved in the University of Edinburgh RDM programme providing support, advice and bespoke training to academics, support staff and research students. We also actively promulgate Edinburgh DataShare's mission by promoting and showcasing new repository developments through institutional, national and international scholarly engagements such as conferences, seminars, workshops, lectures, demonstrations, and the Edinburgh Research Data Blog. Data Library staff are actively involved in the international research data and repository community, including membership, committee work, and office bearing responsibilities in organisations such as the International Association for Social Science Information Service and Technology (IASSIST), Open Repositories (OR), and the Research Data Alliance (RDA). The University of Edinburgh is a member of the Digital Preservation Coalition (DPC), Duraspace, and the Confederation of Open Access Repositories (COAR).

Links to supporting documentation:

- Edinburgh DataShare 'Trustworthiness':  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/trustworthiness>



- University of Edinburgh Research Data Management Policy and Roadmap:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management/rdm-programme>
  
- University of Edinburgh Mission statement: <http://www.ed.ac.uk/about/mission-governance/mission>
  
- Edinburgh DataShare “Preservation policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/preserva>

- Edinburgh DataShare “Service level definition”  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/service-l>
  
- Edinburgh DataShare “Depositor agreement”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/depositor-agreement>
  
- Benefits of Deposit -  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/benefits>
  
- Edinburgh DataShare entry in the Directory of Open Access Repositories (DOAR):  
<http://www.openoar.org/find.php?format=full&rID=1176>

[Last accessed 1 June 2015]

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.**

### *Minimum Required Statement of Compliance:*

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

## **Applicant Entry**

### *Statement of Compliance:*

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

### *Self-assessment statement:*

As part of our quality assurance strategy and as stated in the repository's submission policy, items may only be deposited into Edinburgh DataShare by accredited members of the institution, or their delegated agents. Researchers depositing in Edinburgh DataShare log into the repository with their single sign-on institutional authorisation credentials which are administered by the UK Access Federation, a federated identity management system used by UK Higher Education institutions which authenticates the depositor using Shibboleth technologies.

Upon receipt of a new submission, repository staff check that no personal or sensitive data are included as detailed in statement 1.

Edinburgh DataShare specifies legal requirements in relation to violations of copyright and data protection in the "Depositor Agreement" which the depositor agrees to when they submit their data. A copy of the Depositor Agreement is sent to each depositor via email after deposit. Through this agreement, the depositor grants permission to the repository to immediately withdraw the data if proof of any legal rights violation is received. In case of withdrawal, the repository reserves the right to retain the metadata record and to state that the data has been removed.

The University of Edinburgh adheres to the UK Research Integrity Office "Code of Practice for Research." Researchers depositing in the repository need to comply with this code. In addition, university research centres, schools and colleges have discipline-specific research regulations and ethics committees. These committees specify ethical clearance forms and procedures for research undertaken at departmental, school or college level. Academic research staff must comply with "The Data Protection Act 1998" and ensure that appropriate permissions and rights are cleared prior to data collection and/or public dissemination of research outputs. Rights and permission clearance are addressed in Step-5 of the "Checklist for deposit." Researchers and repository staff can seek advice from Records Management in relation to data protection issues including anonymisation of personal data, confidentiality agreements, and transferring data to other organisations.

Edinburgh DataShare recommends and offers by default the use a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence. Alternatively depositors may manually choose their preferred licence or enter a copyright statement.

Edinburgh DataShare policies and procedures comply with the University of Edinburgh Information Security Policy.

Links to supporting documentation:

- Edinburgh DataShare “Submission policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/submission-policy>
- Edinburgh DataShare “Depositor agreement”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/depositor-agreement>

**Data Seal of Approval Board**

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

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

- UK Research Integrity Office “Code of Practice for Research”:  
<http://www.ukrio.org/publications/code-of-practice-for-research/>
  
- “University of Edinburgh data protection policy”:  
<http://www.ed.ac.uk/schools-departments/records-management-section/data-protection/data-protection-policy>
  
- Edinburgh DataShare “Checklist for deposit”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/checklist>

- University of Edinburgh “Authentication and authorisation”:  
<http://www.ed.ac.uk/schools-departments/information-services/computing/computing-infrastructure/authentication-authorisation>
  
- University of Edinburgh Records Management Section:  
<http://www.ed.ac.uk/schools-departments/records-management-section>
  
- University of Edinburgh “Information security policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/security-policies/security-policy>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## **6. The data repository applies documented processes and procedures for managing data storage.**

*Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Edinburgh DataShare “Service level definition” contains detailed information on the repository’s availability, resilience, backup, and disaster recovery procedures. Hardware and network facilities including storage, and security processes are administered by Information Services IT Infrastructure on behalf of EDINA and Data Library which maintains its own private organisational risk register.

Edinburgh DataShare is hosted in a Solaris container on a larger host (Storage Area Network) which complies with the University of Edinburgh “Security policies.” Containers operate with fewer privileges than their host, making privilege-oriented attacks difficult. The core OS binaries, libraries and kernel modules are immutable (read-only loopback). Access to the repository container is restricted by IP address (SSH) via an application load balancer (with security filters) which can only be accessed from within the University of Edinburgh network domain.

The repository is replicated using Rsync utility software at a separate site (1.5 miles away) every hour. Files are copied in blocks in the replication site, synchronising the directories by checking the time stamp. The mirrored copy is available for failover should the main site become unavailable and would be swapped to the replication site as part of disaster recovery procedures. Backups are conducted daily using IBM TSM licenced backup software on disk and tape (initial backup with permanent increments), and copies are retained for 30 days with a monthly archive copy retained for a year. In the event of data corruption affecting both sites (data corruption or deletion being automatically replicated) or a major external event, a service copy can be restored from either tape backup. Backup infrastructure is geographically separate from the main machine room and the failover site.

Edinburgh DataShare is based on open source repository software (DSpace) originally developed by the Massachusetts Institute of Technology Libraries and Hewlett-Packard, but now supported by Duraspace. The repository developer administers the repository platform according to and in adherence with security models for DSpace servers.

Links to supporting documentation:

- Edinburgh DataShare “Service level definition”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/service-l>
- University of Edinburgh IT Infrastructure Division:  
<http://www.ed.ac.uk/schools-departments/information-services/about/organisation/iti/services>
- University of Edinburgh “Security policies”:  
<http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/security-policies/security-policy>

**Data Seal of Approval Board**

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

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

- DSpace system security: <https://wiki.duraspace.org/display/DSPACE/SecuringDspace>

- DSpace Bitstream Storage Manager:  
<https://wiki.duraspace.org/download/attachments/19006173/BitstreamStorageManager.java?version=3&modificationDate=12901>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

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

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

## 7. The data repository has a plan for long-term preservation of its digital assets.

*Minimum Required Statement of Compliance:*

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

### Applicant Entry

*Statement of Compliance:*

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

*Self-assessment statement:*

The Edinburgh DataShare “Preservation policy” addresses the long-term preservation plans for the repository’s digital assets. Data are retained indefinitely and, where possible, are converted to standard preservation formats to ensure their continued readability at the point of ingest. The repository preserves both the bitstreams of the original data files, as well as those of converted data files. Edinburgh DataShare uses the DataCite service at the British Library to assign digital object identifiers (DOIs) to ensure accessibility of the data. Persistent identifiers are minted when items are deposited, and are included with a suggested citation of a given digital data object.

Failover and backup infrastructure, as detailed in section 6 and in the “Service Level Definition”, ensure the resilience of the repository and long-term bit preservation of digital assets. Data items may be withdrawn from public view for professional or administrative reasons, or if they are found to violate the legal rights of any person, but persistent identifiers and metadata records of withdrawn items remain available online to avoid link rot (citations and publications). In the “Preservation policy” as well as in the “Depositor Agreement” the repository commits to preserve the data and reserves the right to transfer all data to an appropriate repository or archive should the repository infrastructure close down or be superseded.

Edinburgh DataShare offers depositors a list of recommended data formats to ensure long-term preservation of their research data. Depositors are encouraged to submit data in standard preservation formats as mentioned in section 2. Other data formats are accepted, where standard preservation formats are not available or are deemed inappropriate for specific disciplines or research projects. These may be either proprietary and system, software or version dependent or are considered ‘lossy’ (data are lost when compression is applied). Such formats are commonly used within specialised research fields and it is likely the repository will preserve them, but it may not be possible to guarantee the readability of some unusual file formats.

Repository staff regularly monitor technical developments in the field of digital preservation and repository infrastructure, through membership with organisations such as the Digital Preservation Coalition (DPC) and the Digital Curation Centre (DCC) associates network. Repository staff will try to ensure continued readability and accessibility by migrating data items to new file formats where necessary. We are in the process of reviewing retrospective file conversion and normalization workflows, in addition to format identification and migration tools. For example, DSpace curation tasks may be used to systematically complete actions across collections or items in a collection, such as identifying file formats that are at risk of becoming obsolete.

Links to supporting documentation:

- Edinburgh DataShare “Preservation policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/preserva>
  
- Edinburgh DataShare “Service level definition”:  
[http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-l](http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/service-l)
  
- Edinburgh DataShare “Depositor agreement”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/depositor-agreement>

**Data Seal of Approval Board**

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

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

- Edinburgh DataShare “Recommended file formats”:

<https://www.wiki.ed.ac.uk/download/attachments/257254299/fileFormats-d3.pdf>

[Last accessed 1 June 2015]

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 8. Archiving takes place according to explicit work flows across the data life cycle.

*Minimum Required Statement of Compliance:*

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

### Applicant Entry

*Statement of Compliance:*

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

*Self-assessment statement:*

Edinburgh DataShare is integral to the cross-divisional research data management (RDM) programme at the University of Edinburgh. After data have been generated, transformed, analysed, appraised and documented, researchers may choose to deposit in the repository for long-term preservation, publication and re-use. Data preservation workflows are conducted by repository staff, whilst the management of research data prior to their ingestion in the repository are facilitated by a suite of RDM services developed by Information Services to implement the University's RDM Policy and which map onto the research data lifecycle as articulated in the University's RDM Roadmap, such as the 'active data storage' service, DataStore.

The explicit workflows of Edinburgh DataShare are associated with the end of the data lifecycle, and include secure long-term preservation and the sharing and discovery of data assets. When data are submitted to the repository, the preservation workflow starts. The first step is the ingestion of data assets as detailed in "Checking a new item submission", in the DataShare wiki. Depositors are prompted to use the "Checklist for deposit" prior to submitting their data to the repository thereby helping to future-proof their own submission(s).

Once ingestion of data is final, the data are assigned a persistent identifier with an accompanying optional licence and a suggested citation. DSpace is optimized for discovery by search engines and Google Scholar and is OAI-PMH compliant allowing harvesting by aggregation services including Thomson-Reuters Data Citation Index.

Repository staff are key participants in the RDM Programme including data management planning support and consultancy (in particular data preservation, publication and re-use). Researchers seeking grant funding are offered advice and guidance for identifying Edinburgh DataShare as their sustainability solution for data outputs (DataShare and sustainability) as part of research data management planning requirements.

Due to our collaboration with Research Space users of RSpace electronic laboratory notebooks can now archive research data generated in the laboratory directly into the repository using the SWORD protocol to transfer files and metadata. Researchers may do this to time-stamp results from a particular experiment midway through the data lifecycle or when they are finished with the entire project at the end of the data lifecycle.

**Data Seal of Approval Board**

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

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

Links to supporting documentation:

- University of Edinburgh Research Data Management Roadmap:  
<http://www.ed.ac.uk/schools-departments/information-services/about/strategy-planning/rdm-roadmap>
  
- University of Edinburgh “Research Data Management Policy” :  
<http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>
  
- Edinburgh DataShare “Preservation policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/preserva>

**Data Seal of Approval Board**

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

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



- Edinburgh DataShare “Checking a new item submission”:  
<https://www.wiki.ed.ac.uk/display/datashare/Checklist%3A+Checking+a+new+Item+Submission+to+DataShare>
- Edinburgh DataShare “Checklist for deposit”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/checklist>
- DataShare and sustainability: <http://edin.ac/1PHv2sO>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.**

*Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Upon submission of an item into Edinburgh DataShare, depositors agree to the terms and conditions specified in the “Depositor agreement.” By agreeing, the depositor allows the repository to translate, copy, re-arrange or undertake any preservation and curation actions in order to ensure availability and accessibility of digital objects. To ensure access to digital objects, the repository is granted permission to reproduce, transmit, broadcast or display the data in formats or software other than the ones in which the data were originally created or submitted. Although the service takes every care to preserve digital objects and associated metadata, a disclaimer states the repository is not liable for the loss or damage of research data, and takes no responsibility for mistakes, omissions or legal infringements of deposited digital objects. The repository is also granted permission to transfer the data deposited in the current repository infrastructure to another system should Edinburgh DataShare cease to exist.

The procedures that ensure the access and availability of data assets in Edinburgh DataShare are detailed in the “Service Level Definition” under the headings availability, resilience, backup, and disaster recovery.

Items in the repository are open access by default. The repository offers depositors the option of openly licensing their data under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence or to manually add their preferred licence or rights statement. The repository allows a limited-term embargo of data items in which only the metadata record is viewable. Users may request a copy of embargoed data items by clicking on the padlock icon triggering an email to the depositor. Only the depositor may grant access to an individual request during the embargo period. Once the embargo is lifted or expires the data become publicly available.

Links to supporting documentation:

- Edinburgh DataShare “Depositor agreement”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/depositor-agreement>

- Edinburgh DataShare “Service level definition”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/service-l>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

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

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

## **10. The data repository enables the users to discover and use the data and refer to them in a persistent way.**

*Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

The repository offers search facilities (including faceted browsing) to enable discovery of data holdings. Users can search across or within Edinburgh DataShare communities and collections including the full-text of documentation files predetermined for indexing purposes (PDF, txt, docx). Records can be filtered and sorted according to a range of criteria including title, date accessioned, data creator, subject, keyword, funder, spatial coverage. Items in the repository are optimised for discovery in search engines and Google Scholar.

Only items with accompanying documentation are approved for ingest into the repository. Users can view or download documentation and other contextual information to assist re-use. This may include technical reports, codebooks, readme files or software code. Each file can be viewed or downloaded individually or all together as a zipped bundle for convenience. The repository links to the external location of any related publications where they exist, in the full item record.

Expertise relating to the data item resides with the depositor but repository staff will endeavour to assist users with any queries they may have about utilising the data including putting them in touch with data creators themselves. There is a prominent link to contact details on the website banner on all pages.

Email queries are routed to the university call management system where calls are managed and tracked.

Edinburgh DataShare complies with OAI-PMH protocol to enable the indexing and harvesting of the metadata and the online dissemination of data records. The repository also provides a suggested citation based on selected fields in the DCMI metadata schema. The repository uses the British Library DataCite service to mint persistent digital object identifiers (DOIs). The suggested citation plus DOI enables users to properly refer to and cite the dataset in any publication. We believe that this both aids discovery and online access, and provides recognition and reward for data sharing.

Links to supporting documentation:

- Edinburgh DataShare “Metadata schema”: <https://www.wiki.ed.ac.uk/display/datashare/metadata>
- Edinburgh Datashare OAI-PMH compliance: <http://datashare.is.ed.ac.uk/dspace-oai/request?verb=Identify>
- British Library DataCite Service: <https://www.datacite.org/BL>
- Data Library “Contact Us”:  
<http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/contact-us>

[Last accessed 14 July 2015]

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 11. The data repository ensures the integrity of the digital objects and the metadata.

### *Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

#### *Statement of Compliance:*

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

#### *Self-assessment statement:*

Edinburgh DataShare assumes responsibility for preserving and making openly available research data and contextual documentation generated by University researchers in accordance with the Edinburgh DataShare “Preservation Policy.”

The “Checklist for Deposit” and the “New Submission Checklist” details the responsibilities of both depositor and repository staff to ensure that data files are properly prepared and organised and that documentation is accurate and understood. As part of these processes repository staff undertake integrity checking and quality assurance of each deposit (data, documentation and metadata) at the stage of ingest and mediate with depositors where necessary: for example insufficient contextual information to enable re-use; unknown file type; file read error; missing cases or variables in data file with reference to documentation.

In some circumstances metadata may be edited or updated without the need to withdraw an item. However, for any significant changes to metadata or any changes to data files the item will be superseded in favour of a new version to ensure the integrity of the data item.

Edinburgh DataShare uses the DSpace checksum digest algorithm to generate an MD5 checksum for each bitstream (file) which is ingested to ensure the integrity of the digital content from the point of deposit onwards. Fixity checks are executed automatically on a daily basis. The system generates an email alert when changes are detected, so that issues can be addressed immediately and do not get replicated. In addition to the checksum, each file is virus checked upon ingest.

Plans to include checksum values in the item file(s) description for downloaded files are being investigated to allow users to validate data file checksums within their own computing environment. Repository administrators can view the checksum record in the dc:description.provenance metadata field for each bitstream in an item.



Links to supporting documentation:

- Edinburgh DataShare “Metadata schema”: <https://www.wiki.ed.ac.uk/display/datashare/metadata>
  
- Edinburgh DataShare “Preservation policy”: <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/preserva>

- Checklist for Deposit - <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/checklist>
  
- Checklist: [Checking a new Item Submission to DataShare](https://www.wiki.ed.ac.uk/display/datashare/Checklist%3A+Checking+a+new+Item+Submission+to+DataShare) - - <https://www.wiki.ed.ac.uk/display/datashare/Checklist%3A+Checking+a+new+Item+Submission+to+DataShare>
  
- DSpace Checksum Checker - <https://wiki.duraspace.org/display/DSPACE/Configure+checksum+checker/>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

**Data Seal of Approval Board**

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

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

Accept

*Comments:*

## 12. The data repository ensures the authenticity of the digital objects and the metadata.

### *Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

#### *Statement of Compliance:*

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

#### *Self-assessment statement:*

Edinburgh DataShare uses the DataCite service from the British Library to assign digital object identifiers (DOIs) to ensure accessibility and authenticity of the data. Persistent identifiers are minted when items are deposited, and are included with a suggested citation of a given digital object. All items in the repository are open to public scrutiny for the purposes of scientific validation.

As stated in the repository's submission policy, items may only be deposited into Edinburgh DataShare by accredited members of the institution, or their delegated agents. Researchers depositing in Edinburgh DataShare log into the repository with their single sign-on institutional authorisation credentials which are administered by the UK Access Federation, a federated identity management system used by UK Higher Education institutions which authenticates the depositor using Shibboleth technologies.

The submission policy contains a disclaimer that the validity and authenticity of the content of submissions is the sole responsibility of the depositor at the point of ingest. Quality assurance checks by repository staff are detailed in statement 11.

All actions taken on data items within DSpace are logged allowing an audit trail to be examined. Even when a data item is withdrawn a 'tombstone' metadata record remains for purposes of provenance.

The authenticity of digital objects and metadata is protected by the computing environment in which the repository operates. Edinburgh DataShare is hosted in a Solaris zone which operates with fewer privileges than their global zone counterpart. This makes privilege-oriented attacks far more difficult to achieve. Additionally, the core OS binaries, libraries and kernel modules are all effectively immutable in the default configuration since they are provided using read-only loopback mounts from the global zone. Login access to the DataShare zone is restricted by IP address. Access to the DataShare application is via an application load balancer which includes security filters. It cannot be accessed directly from outside Edinburgh University network domain.

### **Data Seal of Approval Board**

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

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

Links to supporting documentation:

- Edinburgh DataShare “Submission policy”:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/submission-policy>

[Last accessed 1 June 2015]

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### **13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.**

*Minimum Required Statement of Compliance:*

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

#### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

We are in the process of reviewing the repository workflows and technical environment with a view to implementing the OAIS reference model. An initial mapping of the OAIS model to our processes and functions has been conducted. A further gap analysis exercise will aim to identify functional archival processes that require attention, rationalize practice, inform policy decision making, and yield a more comprehensive understanding of archival standards that can be used to best deliver efficiency and trusted repository status. Our immediate investigations centre around file format registry, normalisation and migration, and sustainable preservation workflows. We work to review and improve the DSpace architecture as part of the Duraspace open source community, and our own infrastructure as part of the University's RDM Programme and Roadmap activity on an ongoing basis.

Links to supporting documentation:

- University of Edinburgh Research Data Management Programme:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 14. The data consumer complies with access regulations set by the data repository.

### *Minimum Required Statement of Compliance:*

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

## Applicant Entry

### *Statement of Compliance:*

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

### *Self-assessment statement:*

Datasets held in Edinburgh DataShare are non-disclosive, public-use files and can be downloaded without any end-user agreement or restriction. Data consumers (users) are responsible for complying with any licence or rights statement as stipulated by the data producer and conveyed by the repository (see statement 16).

Depending on the agreement that the data creator has with the research funding body or their publication schedule data can be stored in the repository for up to 5 years before it is made openly available. Consumers are thus made aware that anything deposited will eventually become public after any explicit embargo period. When an item is under embargo, the metadata record and file names are visible, but the file contents cannot be viewed or downloaded.

The University of Edinburgh RDM Policy recommends deposit in an “appropriate national or international data service or domain repository, or a University repository.” If a data producer needs to restrict access (e.g. to confidential data) or requires a legal access agreement for a dataset, they would need to use a service other than Edinburgh DataShare.

Links to supporting documentation:

- University of Edinburgh Research Data Management Policy - <http://www.ed.ac.uk/is/research-data-policy>



[Last accessed 1 June 2015]

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.**

*Minimum Required Statement of Compliance:*

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

**Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Datasets held in Edinburgh DataShare are non-disclosive, public-use files and can be downloaded without any end user agreement.

Where applicable Edinburgh DataShare would expect data users to adhere to the UK Research Integrity Office's Code of Practice for Research (2009) as well University UK's Concordat to Support Research Integrity (2012). In addition RCUK Policy and Guidelines (2013) are "intended to apply across the full spectrum of research and training funded by the Research Councils and should be amplified in specific disciplines by the guidance issued by individual Research Councils, other funders, professional associations and learned societies." There would be an expectation that data users from non-UK academic institutions would comply with equivalent research codes of practice for the relevant territorial jurisdiction.

The University of Edinburgh Research Ethics and Integrity Review Group (REIRG) ensure that research integrity and governance has a strong profile at Edinburgh and is firmly embedded in the University's ethos and culture. It ensures compliance with the Universities UK Concordat to support research integrity as well as Funding Council, Research Council and other funders' terms and conditions, ensuring that information on all aspects of integrity, ethics and governance are visible and up to date. REIRG also acts as the University's point of contact with the UK Research Integrity Office and ensures compliance with internal and statutory reporting requirements.

Users from universities, NHS organisations, private sector bodies and charities are subject to UKRIO's Misconduct Investigation Procedure, a step-by-step manual for investigating allegations of fraud and misconduct in research, applicable to all subject areas.

Links to supporting documentation:

- UK Research Integrity Office, Code of Practice for Research (2009): <http://www.ukrio.org/wp-content/uploads/UKRIO-Code-of-Practice-for-Research.pdf>.
  
- Universities UK, The Concordat to Support Research Integrity (2012): <http://www.universitiesuk.ac.uk/highereducation/Documents/2012/TheConcordatToSupportResearchIntegrity.pdf>.
  
- RCUK Policy and Guidelines on Governance of Good Research Conduct (2013): <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/reviews/grc/RCUKPolicyandGuidelinesonGovernanceofGoodResearchPractice.pdf>.

**Data Seal of Approval Board**

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

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

- University of Edinburgh Research Ethics and Integrity Review Group:  
<http://www.ed.ac.uk/schools-departments/governance-strategic-planning/research/research-integrity/reirg>

- UKRIO Misconduct Investigation Procedure:  
<http://www.ukrio.org/publications/misconduct-investigation-procedure/>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

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

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

## **16. The data consumer respects the applicable licences of the data repository regarding the use of the data.**

*Minimum Required Statement of Compliance:*

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

### **Applicant Entry**

*Statement of Compliance:*

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

*Self-assessment statement:*

Datasets held in Edinburgh DataShare are non-disclosive, public-use files and can be downloaded without any end-user agreement or restriction. Data consumers (users) are responsible for complying with any licence or rights statement as stipulated by the data producer and conveyed by the repository.

Data depositors are given two options to licence to their data as part of the ingest process. They can choose from a Creative Commons Attribution 4.0 International (CC BY 4.0) licence recommended by DataShare.

CC-BY 4.0 is written specifically to include databases/datasets as well as creative expressions in all legal jurisdictions. Consumers can copy and redistribute the material in any medium or format, remix, transform, and build upon the material for any purpose, even commercially. The licensor (data producer) cannot revoke these freedoms as long as users follow the licence terms. Consumers must give appropriate credit, provide a link to the licence, and indicate if any changes were made to the data.

Alternatively the data creator can select 'no licence' and fill in their own rights statement. Commercial re-use of full data items and/or metadata is permitted as detailed in the repository's Data and Metadata Policies.

By using a standard open licence, producer and consumer alike may become better informed in relation to norms of data citation and re-use. Repository staff encourage depositors to use an open licence instead of a rights statement due to the difficulty in policing bespoke terms and conditions in an open access repository.

Links to supporting documentation:

- Creative Commons Attribution 4.0 International (CC BY 4.0) licence:  
<http://creativecommons.org/licenses/by/4.0/legalcode> and 'human-readable' summary:  
<http://creativecommons.org/licenses/by/4.0/>

- Edinburgh Datashare Data and Metadata Policies:  
<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policies/data-met>

[Last accessed 1 June 2015]

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

**Data Seal of Approval Board**

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

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

*Comments:*