



Edinburgh DataShare

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:

Institutional repository

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of Repository

Edinburgh DataShare is an institutional repository of research datasets produced at the University of Edinburgh. We accept data from all academic disciplines studied at the institution, up to a reasonable size limit based on technical and resource constraints. However we do encourage our researchers to deposit with disciplinary repositories where that is more appropriate, for example for reasons of discoverability in their research community or functionality more closely tailored to their data. DataShare contains data from a wide variety of disciplines, and therefore a diversity of content types and file formats with 2,950 items ingested and managed at time of submission.

The repository is promoted as an integral and longstanding component of the University's Research Data Service which enables and incentivises its users to share their research data openly as part of good research data management practice. It is operated by the Research Data Support team, part of the Library & University Collections (L&UC) division of the University, with developer input from the Digital Library team and EDINA. The repository has been running continuously since its inception in 2008, including during a reorganisation involving the transfer of core staff from EDINA to L&UC in 2018. It was certified for the Data Seal of Approval in 2015. The service owner set a key objective to apply for Core Trust Seal certification during the current Research Data Service Roadmap three-year period, which has been ratified by the service's steering group.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept
Spell out EDINA.

Brief Description of the Repository's Designated Community.

Our Designated Community encompasses research staff and research students of the University of Edinburgh and academic and independent researchers from outside the University. Our end-users include researchers of any academic discipline, independent and public sector and NGO/charity sector researchers worldwide. Our depositors are the staff and

research students of the University of Edinburgh, from our three Colleges and twenty different Schools, covering an extremely wide range of all the main traditional and modern academic disciplines of the arts and humanities, social sciences, physical sciences and medical science. While our depositors are not necessarily (as individuals) consumers of the data, many of them will be end-users (downloading their own and their colleagues' data for reference and re-use) and they will be potential re-users. Therefore we regard our University of Edinburgh community of depositors and support service users as a useful microcosm of the global academic/research community. We view all our staff and students as not of a specific academic discipline, rather it is any researcher, from any community that can put any research datasets together, ie our support is not domain-specific.

Some datasets are of interest to the wider public, especially those discussed in the press, social media or linked to from other websites such as Wikipedia and Wikidata.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

All submissions undergo basic curation as per level B.

We do apply the standard that the research data should be re-useable, for the purposes of research, if that is at all possible. We put onus on the depositor to deposit the data in a form that is sustainable. Thus the significant properties which are relevant in this context are evaluated by the depositor primarily, with our guidance, and by the DataShare curators.

Occasionally where a curator identifies a more suitable file format is available, submitters are encouraged to convert their

own files before a deposit is accepted. All data items are required to include some form of accompanying documentation. Our depositor agreement does allow for us to convert to new formats after deposit, as our preservation planning indicates is advisable, and which we have done in a number of deposits. When we convert any files, we manually check them afterwards for readability and internal functions such as formulae. Whereas for example look and feel would not be considered significant properties in this context, it is the information content that is relevant (therefore we don't need to preserve for look and feel in the way that some other archives would). We always preserve the original bitstream, so if we ever discovered a formula that had broken, we could go back to the earlier file.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Outsource Partners. If applicable, please list them.

N/A

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept

Other Relevant Information.

The University Library is a "silver" member of DSpace, in order to contribute to the Open Source community from which we benefit, in the use of their repository software. <https://duraspace.org/dspace/community/members>

Edinburgh DataShare is listed on re3data at <https://www.re3data.org/repository/r3d100000047> and indexed in the Data Citation Index (<https://clarivate.com/webofsciencelibrary/solutions/webofscience-data-citation-index/>). (A search indicates ~50 citation records at the present time.) We participate in the IRUS-UK COUNTER-conformant usage statistics repository aggregator service (<https://irus.jisc.ac.uk>). We are a member of the British Library UK consortium for DataCite and the University of Edinburgh is a subscriber to Altmetric and a member of ORCID.

Significant changes since our last application

Since the last application, we have enhanced the service to permit larger datasets, up to 100 GB, and we have added many features:

- a REST API has been activated and is publicly available;
- a request-a-copy facility for those deposits under temporary embargo;
- drag'n'drop file upload;
- creation of tombstone records for deposits rather than withdrawing them (for example where an intellectual property conflict has been identified), so that the metadata remains publicly available, even though the files are blocked. - for embargoed deposits, the system sends an email notification to the depositor one week before the expiry of the embargo;
- added an Altmetric badge;
- upgraded the DSpace system to version 6.

We have also expanded our base of depositors to an even wider range of departments across the University.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Revision accepted.

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
Accept

Reviewer 2

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

Response:

“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.”

Trustworthy Digital Repository web page:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/trustworthy-digital-repository>

This page was added by the service owner following DSA certification in 2015 and updated in 2019. It goes on to explain the importance of preservation to our research community:

“Researchers, the repository and trust:

Perhaps even more important than attaining an official standard is knowing that researchers trust Edinburgh DataShare. What makes data producers more likely to share their data using our repository? We know that items in the repository score high in search results on Google and Google Scholar.

We also emphasise the benefits for depositors down the line, for example when they move to another institution, to be able to easily obtain their data, knowing that its integrity is secured.

What makes users more likely to use datasets they find in our repository? We hope that the careful guidance we give data producers about how to make their data usable, the quality checks we do upon submission, as well as the rule that every data item must include some human-readable documentation, reaffirms our commitment and reassures users that they are accessing and using quality data.

We also hope that the University of Edinburgh’s reputation for world-class research lets them know that they are getting authentic, state of the art research data outputs from the DataShare repository.”

The University of Edinburgh supports the mission of the repository to make data accessible for the long-term, at the highest level, as evidenced by the institutional Research Data Management Policy, passed by the University Court in May, 2011:

“The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.[...]

“Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.”

<https://www.ed.ac.uk/is/research-data-policy>

We believe the University’s own mission statement is also relevant:

“The mission of our University is the creation, dissemination and curation of knowledge.”

<https://www.ed.ac.uk/governance-strategic-planning/content-to-be-reused/vision-and-mission>

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Accept

"Researchers, the repository and trust" seems a title. Consider "Researchers, the Repository and Trust" or "Researchers, the repository and trust:"

II. Licenses

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

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Overview – Procedural and policy measures

The DataShare submission form offers users a Creative Commons attribution licence as the default ('CC-BY'), with the option to instead make their own statement concerning rights. As documented in our Curator Checklist (referenced below), whenever the user does not select the default licence, the curator always carefully checks the rights statement they have instead made, and will check with the depositor that they have ensured the appropriateness of their specified conditions of access in the context of any requirements of their funder(s) and the University of Edinburgh Research Data Policy. Additionally, curators always remind depositors intending to apply extra conditions on re-use that the service does

not guarantee end-users' compliance with the depositor's conditions of access, since it would not be practicable for the service to pro-actively monitor the whole internet for non-compliant re-use of data. We advise depositors if they need to restrict re-use of the data, and it is appropriate to do so, they should restrict access, e.g. by depositing in our restricted archive (Edinburgh DataVault) instead of DataShare.

Default licence

Early depositors were offered the Open Data Commons ODC-BY attribution licence by default.

EXAMPLE: MacCallum, Stuart N; Merchant, Christopher J. (2011). ARC-Lake v1.1 - Per-Lake, 1995-2009 [Dataset]. University of Edinburgh. School of GeoSciences / European Space Agency.
<https://datashare.ed.ac.uk/handle/10283/130?show=full>

Where the current default licence applies, we put "Creative Commons Attribution 4.0 International Public License" in the dc.rights field.

Although we do not have a published non-compliance statement, on those deposits with the default licence (the vast majority) the licence is always displayed prominently on the deposit landing page (summary view) and is included in the dataset as a file, to bring it to the attention of the end user.

EXAMPLE: Recent deposit with default licence: Owolabi, Alíz TY; Reece, Sarah E; Schneider, Petra. (2021). Daily rhythms of both host and parasite affect antimalarial drug efficacy, [dataset]. University of Edinburgh. School of Biological Sciences.
<https://datashare.ed.ac.uk/handle/10283/3858?show=full>

Deposits with licenses other than the default

On those Items where the depositor did not choose the default licence, there is always text in the dc.rights field, checked by the curator, either specifying a different licence, or a statement by the depositor setting out the equivalent information. Our curators always strongly encourage users to choose the most open licence permitted by their funder or the university. Several examples are provided below.

Restrictions imposed by the data provider

Some restrictions on re-use are imposed by the source of the data.

EXAMPLE: Non-commercial licence rights statement: Paretì, Paolo; Klein, Ewan H.. (2016). The Human Know-How Dataset, 2014 [dataset]. <https://datashare.ed.ac.uk/handle/10283/1985?show=full>

Other Creative Commons licences

Some deposits use other Creative Commons licenses for example some artistic works where the depositor has placed a 'non-commercial or non-derivative re-use only' restriction on the data.

EXAMPLE: Wilson, Graeme B; MacDonald, Raymond AR; Duff, Malcy; Duncanson, Lindsay; Gabrysch, Marek; Green, Owen; Keay, Cath; Knox, Dawn Felicia; Lloyd, Emma; Lozano-Thornton, Dario; Masson, Sheila; Parr-Burman, Michael; Robertson, Ali; Stockbridge, Jamie; Smith, Grant; Wallder, Dan; Williams, Sean. (2015). Meanings of Interaction Among Musicians Improvising (MIAMI), 2014 [sound]. University of Edinburgh. Edinburgh College of Art. Reid School of Music. <https://datashare.ed.ac.uk/handle/10283/830?show=full>

Public Domain

A small number of deposits are released under a Public Domain Dedication Licence. This may be beneficial in respect of re-use of data in Wikimedia Commons, Wikidata and Wikipedia.

EXAMPLE: Public domain data: Blackwood, Carol. (2017). GB Roman Roads, [Dataset]. EDINA.

<https://datashare.ed.ac.uk/handle/10283/2427?show=full>

Anonymised data

Some deposits of anonymised clinical trial data, where there is only a very remote risk of disclosure or harm, are kept under embargo, with selective access, conditional on end-users signing a Data Access Agreement promising to use the data only for academic research and not to attempt to identify participants. The Data Access Agreement is administered by a local staff member in the research unit.

HISTORIC EXAMPLE (embargo now expired): Sandercock, P; Wardlaw, J; Lindley, R; Cohen, G; Whiteley, W. (2016). The third International Stroke Trial (IST-3), 2000-2015 [dataset]. University of Edinburgh & Edinburgh Clinical Trials Unit. <https://doi.org/10.7488/ds/1350>

All rights reserved

A very few deposits are labelled 'All rights reserved' – these depositors have been made aware that the University cannot necessarily prevent infringements.

EXAMPLE: Brown, Nick. (2018). ePython - a tiny Python implementation for micro-core architectures, [software]. University of Edinburgh. EPCC.

<https://datashare.ed.ac.uk/handle/10283/3072?show=full>

Software

A small number of software deposits have GNU or MIT licenses; if such depositors are unsure about their choice of licence, the curation team refers them to our local colleagues at the Software Sustainability Institute (SSI) for expert advice.

EXAMPLE: Goddard, Benjamin. (2017). Wave packet dynamics in the optimal superadiabatic approximation - Matlab code, [software]. <https://datashare.ed.ac.uk/handle/10283/2645?show=full>

References

Curator 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>

Reviewer Entry

Reviewer 1

Comments:

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

The amount of information provided to data producers about file formats suitable for preservation and service availability is very detailed and convincing.

I also noticed the information you provided to prove that repository's content will be take over by another institution in case of closure of the service.

However, to achieve compliance level 4, a formal, written agreement between the repository or the repository's parent institution and another organisation, that will take over responsibility in case of a shutdown, is necessary.

Reviewer 2

Comments:

3 – The repository is in the implementation phase

Response:

The Edinburgh DataShare "Preservation policy" addresses the long-term preservation plans for the repository's digital assets and exit strategy, in the unlikely event of the closure of the service, through migration to another non-disciplinary repository.

The University's vast infrastructure, modern data centre facilities and specialised information services teams (hundreds of

staff, co-ordinating support activity across teams with a call-management system), support the very high level of availability of the DataShare service, meaning that the risk is very low, of unplanned or extended disruptions through technical issues (Rice et al. 2017). Furthermore, interruptions are kept very short as a result of the support available from on-site university staff. The DataShare service is run in accordance with the university's strategy for service delivery, which encompasses incident management, problem management, change management and release management, all geared to maximise availability. For example, every time a release is scheduled, it goes through an approval process; major changes are scrutinised by a Change Advisory Board, and evidence of a roll-back plan is provided. Several members of the team including the service owner have qualifications in the ITIL Service Management framework. See under the 'Availability' sub-heading in the DataShare 'Service level definition', link at the end of this response.

Business continuity planning is done at an infrastructure level, by the team responsible for the main University data centre, and the Library has done business continuity planning in terms of staff availability to work from home in an emergency. These have been 'tested' to satisfaction during the Covid-19 pandemic, with all DataShare staff working from home for many months, with no disruption caused to the service.

The University of Edinburgh Library has been operating for over 400 years, and digital curation has become a core element of its strategy. If the university ever decided to shut down DataShare, there would be an exit strategy which would allow for moving content to University of Edinburgh collections/archives and if necessary negotiations with appropriate services such as the British Library, National Library of Scotland or a suitable alternative. The data would be stored and protected while these steps were taken. The Library & University Collections division is substantial and well-resourced, and would have enough resource to support appropriate or necessary interim arrangements.

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.

Repository staff have recently completed a 3-year Research Data Management roadmap, including data stewardship objectives. Strategic direction is approved and monitored by our Research Data Service Steering Group which meets quarterly, and is funded through a Digital Research Services University programme.

Failover and backup infrastructure, as detailed in the "Service Level Definition", ensure the resilience of the repository and long-term bit preservation of digital assets, including disaster recovery. A snapshot copy of the repository is also being stored in our long-term preservation system, Edinburgh DataVault. The vault has already been set up. A test version of this, containing an exported snapshot of the entire repository database (the metadata) and its contents (the assetstore ie the files) has been stored. Our tech team is working on creating an AIP / METS version using a standard function of DSpace. Updates will be added annually.

Data items may be withdrawn from public view for professional or administrative reasons, or if they are found to violate the

legal rights of a particular person, but persistent identifiers and metadata records of withdrawn items remain available online to avoid link rot (in 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. 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’ (i.e. granularity is 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 more unusual file formats.

The University policy is backed with substantial amount of resource invested over the years as per the institution’s RDM Roadmap:

<https://www.ed.ac.uk/information-services/about/strategy-planning/rdm-roadmap>

Preservation policy

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/preservation-policy>

Service level definition

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/service-level-definition>

Choose the best file formats

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats>

Rice, R; Fergusson, D; (2017) Research Data Management at the University of Edinburgh: How is it done, what does it cost? In: LEARN Toolkit of Best Practice for Research Data Management. (pp. 91-94). Leaders Activating Research Networks (LEARN). <https://doi.org/10.14324/000.learn.18>

Reviewer Entry

Reviewer 1

Comments:

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
Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

All data submissions are checked by curators before being accepted into the repository and made open. The curators are trained to spot potential infringement of data protection regulations.

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 staff members of the institution, their bona fide collaborators, postgraduate research students or undergraduate students depositing with permission from their supervisor.

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 the Shibboleth technology.

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, and a copy of the Depositor Agreement is stored as part of the item in DSpace. 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 have 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. We do not accept unconsented personal and sensitive data. We have quality assurance procedures to check new submissions for personal data. On occasion we will query depositors in respect of any data which appear to be actually or potentially disclosive. This might involve requesting removal of some potentially disclosive details. On occasion prospective depositors, following our advice, will decide to store their data in our restricted system (Edinburgh DataVault) instead of DataShare. Where there is anonymised data with a low level of risk of re-identification and harm, we allow DataShare depositors to use the embargo feature over a multi-year period to facilitate selective access. The curation team benefits from access to training and advice from the University of Edinburgh Data Protection Officer and her staff. Members of the curation team have undertaken specific training on disclosure risk.

University of Edinburgh Data Protection guidance

<https://www.ed.ac.uk/data-protection/data-protection-guidance/specialised-guidance/research-data-protection>

Takedown Policy

DataShare Depositor Agreement (see “Ownership” re copyright; and see “Removal of the dataset from Edinburgh DataShare” re takedown policy).

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/depositor-agreement>

DataShare Preservation Policy (see “6. Acceptable reasons for withdrawal”)

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/preservation-policy>

University of Edinburgh – Research Integrity – Our commitment

<https://www.ed.ac.uk/research-office/research-integrity/our-commitment>

DataShare – Checklist for deposit

Step 5 – Permissions and Rights: “Have you sufficiently anonymised your data, or obtained explicit consent from any data subjects whose identity could be revealed from the data (including images)?”

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>

DataShare – Service Level Definition (lists users eligible to submit data for deposit)

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/service-level-definition>

Reviewer Entry

Reviewer 1

Comments:

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
Accept

Reviewer 2

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

Response:

The Research Data Support function of Information Services, University of Edinburgh, was brought together from two teams – one in EDINA and one in Library & University Collections – in 2018. This was a rationalisation in part because of the logic of operating the institutional data repository from the academic library, in keeping with peer institutions. The Edinburgh University Library dates back to 1580, three years before the founding of the University (https://en.wikipedia.org/wiki/Edinburgh_University_Library) and is particularly proud of its archival holdings and special collections, preserved over centuries (<https://www.ed.ac.uk/information-services/library-museum-gallery/crc/collections/archives>), and more recently of its digital collections and digital preservation abilities, managed by the Digital Library team.

Funding for Edinburgh DataShare and its staffing is an integral part of the regular recurring funding of the University's Information Services. The Research Data Service, of which Edinburgh DataShare is a part, was moved onto recurrent funding in 2016. Although other aspects of the service are expected to recover their costs, DataShare is offered free at point of use as an integral research service.

The repository is operated by the Research Data Support team (<https://www.ed.ac.uk/information-services/research-support/research-data-service/contact>) led by Robin Rice (the Data

Librarian and Head, Research Data Support), who also acts as Service Owner of the Research Data Service (of which the repository is a key component). The repository is based within the Library Research Support section along with Scholarly Communications and Research Information Systems. The Digital Library section's Development and Systems team provides technical infrastructure and ongoing development and technical support for the repository. See Library and University Collections Organisational Chart, at <https://www.ed.ac.uk/information-services/about/organisation/structure>.

Four members of staff (approx 2 FTEs), entitled Research Data Support Assistants and Research Data Stewards, do the day to day administration of the repository and liaison with users. They all have research backgrounds, e.g. subject Masters and PhD in complementary disciplines representative of the Designated Community. The first set of duties in the Research Data Support Assistants' job descriptions, updated with the 2018 reorganisation, is "Review data submissions to the data repository (Edinburgh DataShare) according to policy and procedures; ensure adequate metadata and documentation is provided; engage with depositors regarding setup of data collections and links with publications. Organise batch imports for large deposits as needed. Enter records into the data asset register (Pure)," making up about 30% of their time. (Since the time of writing the records are entered into Pure in an automated way, but still need to be approved.) Staff members have an annual development review with their line manager to determine personal goals aligned with organisational strategy and to identify opportunities for professional development, including digital skills courses, regional meetings and national and international conferences (such as Repository Fringe and Open Repositories).

The two Research Data Support Assistants and the Data Librarian are permanent members of staff paid by core University funds distributed to Information Services. Meanwhile the two Research Data Stewards are part-time postgrad student interns, brought in to ensure continuity of the service during a period of relative short-staffing caused by the Covid-19 pandemic. They have been trained to follow all our existing procedures and apply DataShare's high standards in exactly the same way as the Research Data Support Assistants. Further recruitment is planned for later this year. These staff spend a significant portion of their time moderating repository deposits as part of their job descriptions.

<https://www.ed.ac.uk/information-services/research-support/research-data-service/contact>

The system is supported by two internal software developers, both of whom also carry out the sysadmin role and liaise with the support team to maintain, upgrade, develop and repair the repository. One of the developers is based in the University of Edinburgh Digital Library team in L&UC, and the Library is buying out time from the other, a permanent developer in EDINA, to assist with this work, such as the major DSpace upgrade to DSpace 6 in early 2021, and the development work continues to be managed by the Digital Library section. EDINA is a digital solutions provider, hosted at the University of Edinburgh. EDINA is not an abbreviation, it is derived from the old name for Edinburgh 'Edina'. Fortnightly meetings between support staff and development staff drive continual service improvements (bug fixes and new features) on the repository. High level objectives, such as aligning with Open Science by adding a REST API and embedding Altmetrics, are documented in the Research Data Service Roadmap, a rolling three year 'living' document managed by the service owner and approved by the academic-led Research Data Service Steering Group: <https://www.ed.ac.uk/information-services/about/strategy-planning/rdm-roadmap> .

The Library's relevant institutional memberships include Scottish Digital Library Consortium, Confederation of Open Access Repositories (COAR), 'silver' membership of Lyris/Dspace, and the British Library's UK Consortium membership of DataCite.

Rice, R; Fergusson, D; (2017) Research Data Management at the University of Edinburgh: How is it done, what does it cost? In: LEARN Toolkit of Best Practice for Research Data Management. (pp. 91-94). Leaders Activating Research Networks (LEARN). <https://doi.org/10.14324/000.learn.18>

Reviewer Entry

Reviewer 1

Comments:

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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository staff engage actively with different pools of external expertise as follows:

- With the repository community through active participation in the DSpace Community Advisory Team and annual

participation in the Open Repositories international conference and community email lists and social media;

- With the Research Data Management and digital curation community through informal contacts with local colleagues in the Digital Curation Centre and a variety of conferences (IASSIST, RepoFringe, IDCC) events and social media contacts;
- With the digital preservation community through membership of the Digital Preservation Coalition, and attendance at workshops, conferences and training.
- We consult PRONOM and the Library of Congress for the latest information on data types and so on.

We draw upon substantial in-house expertise in digital preservation:

- The full-time Digital Archivist to the university and the Head of Digital Library both provide digital preservation guidance and substantial expertise and are very generous with their time;
- Repository staff participate in the university's Digital Preservation Working Group, convened by the Digital Archivist, which provides opportunities to learn about digital preservation approaches and technology being used across several teams and many projects such as Archivematica.

We communicate with our Designated Community (both internal and external) through a variety of channels:

- For governance purposes, formally, through our Steering Group;
- The Research Data Support team delivers training on our services, and Research Data Management in general, across the institution, reaching staff and research students of all disciplines and roles. We gather feedback through the interactions with users at training.
- Informally through institutional networks, social media and our departmental blog. The Research Data Support team has a dedicated Communications Plan which focuses on key stakeholders within and beyond the University.
- The team's "Research Data Blog" features numerous articles on DataShare, highlighting the benefits of deposit and new features:

<http://datablog.is.ed.ac.uk/tag/datashare/>

- University of Edinburgh Research Data Support twitter account occasionally highlights new deposits and encourages colleagues to deposit data:

<https://twitter.com/ResearchDataUoE>

- On a day-to-day basis, we pick up feedback through our dialogue with submitters throughout the curation process, captured in the in-house service desk software so that expertise from colleagues may be accessed efficiently. We can draw upon expertise within the university, work with the depositor, their colleagues, ask the digital preservation community. We also bring to bear our vast experience of working with research data: very diverse file formats, structures and standards from the very wide range of disciplines from which our new deposits come. We don't rely purely on file extensions, but consult a variety of sources of expertise, as set out in more detail in our response to R10. We have occasionally helped create community preservation knowledge by sending PRONOM information about the formats unknown to them which we have.

Reviewer Entry

Reviewer 1

Comments:

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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The integrity and authenticity of data and metadata are guaranteed by a carefully controlled process encompassing ingest, storage and management of metadata.

Authenticity: Only users with University of Edinburgh login credentials may submit data directly to the repository. On the rare occasions that external users submit data generated from a collaboration with University of Edinburgh researchers, we check their credentials before accepting the data for ingest.

Our Submission Policy makes clear that the authenticity of the data is the responsibility of the depositor.

All submissions are manually checked by our curators, with a workflow enforced by the DSpace repository. Curators inspect all metadata fields, and will open data files to ensure their integrity. Curators ensure the documentation and metadata are (as far as is possible to determine) clear, coherent and sufficiently detailed as to allow for the data to be truly Findable, Accessible, Interoperable and Reusable (FAIR). In the unlikely event that any inauthentic (fake) data were submitted, we are confident that our curators would spot them at this stage. More commonly, we have occasionally found images, logos, or other copyrighted material that a depositor has naively submitted with their data for some reason, which

we have been able to catch and reject, while educating the researcher about copyright permissions.

In addition, the curators assess and decide which DSpace Collection a new deposit should be assigned to – these Collections are closely tied to University organisational units (Colleges and Schools), providing an informal check on the depositor's association with the University. A further check on the data creator's identity takes place retrospectively, when the metadata are copied over to our Pure Current Research Information System (CRIS) and linked to the data creator's profile(s), in a step which is checked manually by our curators.

Integrity: The DataShare DSpace system stores provenance metadata which is visible to curators. This provides an audit trail, which logs the following:

- the identity of the submitter,
- the filenames, sizes and checksums of all files in the submission,
- date of submission,
- a record of whether the submission was rejected and on what grounds,
- date of re-submission,
- the identity of the approving curator,
- details of the files (filenames, sizes and checksums) at the time of approval.

NB Since depositors or curators carry out any file migrations as part of the preparation before the ingest of the data, the action is not formally included in the audit trail. Pre-ingest activity is not audited.

Our DSpace system does not allow for automatic logging of amendments to metadata and data. However, only the highly trained curators have permissions to amend metadata or bitstreams in the repository, so only those amendments consistent with adherence to our standards are allowed. There is an audit trail for such amendments in the University's enquiry handling software (TopDesk).

We frequently decline requests from users to modify data files. We always enforce a version control protocol: that any substantive change to the data requires a completely new dataset with a separate persistent identifier, and detailed annotation on the earlier version to mark it as having been superseded, with reciprocal links in the standard Dublin Core fields. In some instances, such as very new deposits, we will allow very minor changes to metadata or documentation, such as spelling mistakes. In such cases, we pre-pend the title of the earlier version(s) of the dataset with the word 'superseded'. We document the relationship between the different versions using the Dublin Core fields: Supersedes (aka dc.relation.replaces) and 'Superseded by' (aka dc.relation.isreplacedby) to include reciprocal persistent identifiers in their metadata.

EXAMPLE WITH COLLECTION PAGE LINK: Scorza, Livia; Bernotas, Gytis; McCormick, Alistair. (2018). SUPERSEDED - Photometric stereo dataset with annotated leaf masks, [dataset]. University of Edinburgh. School of Biological Sciences. Institute of Molecular Plant Sciences. <https://doi.org/10.7488/ds/2455>.

Any user looking at the landing page of the Collection which contains the two versions of the above dataset can see at a glance that one version is superseded while the other is not:

“PS-Plant: Photometric Stereo data sets”

<https://datashare.ed.ac.uk/handle/10283/3199>

RECENT EXAMPLE: Wilkinson, Tim; Schnier, Christian. (2020). SUPERSEDED - Medication-wide association study of dementia, [dataset]. University of Edinburgh. Usher Institute.

<https://datashare.ed.ac.uk/handle/10283/3581?show=full>

EXAMPLE SHOWING VALUES FOR BOTH ‘Supersedes’ and ‘Superseded By’ fields: SUPERSEDED - Jones, RA; Gillingwater, TH. (2016). NMJ-morph, [interactive resource]. University of Edinburgh. Centre for Integrative Physiology.

<https://datashare.ed.ac.uk/handle/10283/1907?show=full>

In the case of a dataset being superseded, the complete new deposit contains all the files and metadata. The earlier version usually remains available in full. Therefore there is no need for DataShare or the curators to carry out a comparison of the versions. In a small number of cases, we have agreed with the depositor that continued access to the files would be likely to cause confusion among users, in which case access to files is blocked to the public, while the metadata of the earlier version nonetheless remains available ie a tombstone record of the dataset is still available.

EXAMPLE of a superseded, tombstoned deposit: SUPERSEDED - Roy, Marcia; Sorokina, Oksana; McLean, Colin; Gonzalez, Silvia Tapia; DeFelipe, Javier; Armstrong, J Douglas; Grant, Seth. (2018). A synapse proteome resource for mouse brain regions, [dataset]. Edinburgh University. Centre for Clinical Brain Sciences. <https://doi.org/10.7488/ds/2398>.

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 weekly 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.

Checksum values are now displayed for all files, 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.

In addition, we now store an extra copy of each file external to DSpace in order to allow us to provide a download feature that has capacity to allow downloads of larger files (up to 100 GB). Users access this via the ‘Download all’ button and receive a zipped bundle of all the item’s files.

The Edinburgh DataShare repository contents are backed up to tape daily and database snapshots are available on disk for two days. Database snapshots are also backed up to tape. Daily tape backups are retained for 30 days and a monthly archive copy is kept for a year. In addition, a snapshot of the entire repository is in the process of being archived in a new

vault in the Edinburgh DataVault.

Users are expected to provide well-organised and accurate documentation, see step 3 on data preparation and step 4 on the preparation of documentation:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>

Our online tutorial video encourages users to make the data in their submission FAIR:

How to archive your data on Edinburgh DataShare, the open research data repository

https://media.ed.ac.uk/media/How+to+archive+your+data+on+Edinburgh+DataShare/1_9dskcdbi

Curators carry out a rigorous set of checks on all submissions, checking data integrity and metadata coherence according to protocols we document on the DataShare wiki:

<https://www.wiki.ed.ac.uk/display/datashare/Checklist%3A+Checking+a+new+Item+Submission+to+DataShare>

Reviewer Entry

Reviewer 1

Comments:

Examples for and additional information about audit trails are convincing. Plus, the FAIR online tutorial not only provides strong proof of your descriptions regarding your deposit processes, it was also well done and very informative

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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository's Submission Policy states that "Items may only be deposited by accredited members of the institution, or their delegated agents." This means that all academic staff and their external collaborators if they so wish, may deposit data from their research projects in DataShare. This not only fulfils the University Research Data Management Policy by giving researchers a way to properly make their research data available as their project comes to a close, but it provides a quality check on content in line with the level of quality that is expected of University of Edinburgh academic staff and their funded and non-funded research. When it comes to student data, the curators may look for indications that deposition of the data has been recommended by research students' supervisors, which would also provide some level of quality assurance. In practice it is rare for non-research Masters students and undergraduates to attempt to deposit data, because promotion of the service is not aimed at them. Occasionally support staff deposit data from professional projects that may be useful to researchers, or they may deposit legacy research data for which they have come to be responsible. Corporate data is not generally in scope and is disseminated through other means.

Therefore a deposit would generally not be rejected for appraisal reasons, but it might be rejected for other reasons, including not being recognised as research data (not being in scope). We have at times recommended other repositories, such as Edinburgh Research Archive, for ephemeral material such as working papers, conference presentations or posters, or Gitlab/Github for software (though we do accept software outputs that are suitable for the repository, such as models or original code that processes research data and is in a static form). According to the Submission Policy, validity, authenticity and IPR are the responsibility of the depositor, but the curators check for these as is practicable.

Collections are organised according to their alignment with university departments (Colleges, Schools and usually research units and research groups with a named leader or research themes). These organisational structures can change over time and collections can be renamed or moved from one community or sub-community to another when required, without affecting persistent identifiers. Items may be browsed by drilling down from the top level of the three colleges, "Arts, Humanities & Social Sciences", "Medicine and Veterinary Medicine, and "Science and Engineering," or from the support services category.

As the repository is designed to be an internet-facing system, only material that can be made open access, now or in the future are deemed appropriate; so it is important that only non-sensitive material be deposited. Any commercial-in-confidence data or non-consented and non-anonymised personal data must be rejected. Repository staff have been trained in GDPR generally and GDPR for researchers, so they all know how to recognise personal data. Data that appear to be disclosive to individuals would be queried with the depositor, who must clarify that suitable anonymisation is in place or add proof of consent for making it available.

We have discussed our file format registry and 'future-proofing' guidance to depositors in other responses. In this context, where a deposited file does not exist in the repository's file format registry, it is likely to be queried by the curator. If the depositor successfully argues that there is no relevant known standard, or that is the de facto standard in their field, then the curator may choose to accept it, and update the file format registry accordingly. We have also discussed metadata and documentation and how the curators may request richer content before accepting a deposit. The Checklist for Deposit is one example of how the service guides the depositors towards providing better metadata. The training we deliver on good practice in Research Data Management also raises awareness of the importance of good quality metadata, and making

data FAIR.

The text of the Depositor Agreement sets out under which conditions an item may be withdrawn from the repository (such as a breach, complaint or copyright violation). In any such cases a metadata record without files may be kept for reasons of provenance (a tombstone record).

- Submission Policy:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/submission-policy>

- Choose the best file formats:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats>

- Checklist for Deposit:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>

- Depositor Agreement:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/depositor-agreement>

- "Research Data Training: Semester Two, 2020/21"

Courses on the benefits of good practice in research data management, and self-directed learning course MANTRA, raise awareness of FAIR principles:

<http://datablog.is.ed.ac.uk/2021/03/17/research-data-training-semester-2-2020-21>

- "How to archive your data on Edinburgh DataShare, the open research data repository"

This video tutorial demonstrates how the service guides depositors to provide better metadata.

https://media.ed.ac.uk/media/How+to+archive+your+data+on+Edinburgh+DataShare/1_9dskcdbi

Reviewer Entry

Reviewer 1

Comments:

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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The Research Data Service provides detailed documentation of our processes and procedures in managing the archival storage of the data. For our depositors and end-users, we have clear and thorough, user-facing web pages, quality checked and reviewed annually. For the Research Data Support team and technical team, our wiki pages detail all our procedures and are updated as required by team members. We use an issue tracking system to retain further technical information about DataShare's repository system, associated software and the environment.

Our system can be understood in OAIS terms as follows:

- Our batch import process ingests data in DSpace Simple Archive Format, which can be considered a SIP (Submission Information Package).
- The alternative route, the more popular web submission form, guides the user to provide all the same elements of data and metadata, thus creating a virtual SIP, which curators then examine using the DSpace workflow.
- On approval, the new deposit is stored in DSpace, bitstreams, descriptive metadata and administrative metadata, the same elements as found in an AIP (Archival Information Package).
- Now that the University has a long-term golden copy storage facility, the Edinburgh DataVault, we will carry out an annual export of the repository contents to AIP format for storage in the DataVault, such that a copy will be stored in the cloud, and could be used for disaster recovery. Our tech team is actively working on the export, which is a relatively straightforward DSpace function which exports a METS package. The vault structure has been created and configured. An initial test version (a simplified fileset of our test repository) is already stored in the test instance of the DataVault.
- DataShare uses DSpace's authentication tools to provide restricted submission privileges, administered by the Research Data Support team, such that only users already manually added to the Submitters group of a given Collection by a curator may make a submission to that particular Collection.
- DataShare curators from the Research Data Support team use DSpace's data management tools including batch metadata editing and editing of metadata on individual Items.
- The DataShare web interface makes full use of DSpace's Search Engine Optimization features to ensure our metadata may be harvested, and they are harvested by OAIS-PMH, Google Scholar, Google DataSets and so on.

The Edinburgh DataShare repository contents are backed up to tape daily and database snapshots are available on disk for two days. Database snapshots are also backed up to tape. Daily tape backups are retained for 30 days and a monthly

archive copy is kept for a year.

Duplicate deposits almost never occur, given the manual deposit and manual curation process.

We store an extra copy of each file external to DSpace in order to allow us to provide a download feature that has capacity to allow downloads of larger files (up to 100 GB). Users access this via the 'Download all' button.

As described in our response to R7, curators follow a standard process for ingesting, checking and approving data consistently. All new members of the curation team undergo detailed training and supervision in the application of this procedure.

- Edinburgh DataVault

<http://www.ed.ac.uk/is/research-support/datavault>

- DataShare user-facing documentation on the University of Edinburgh website includes pages on preparing to deposit data, service policies and the Depositor Agreement, as well as definitions, benefits of depositing, and the nature of a trusted digital repository:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository>

- Further curator procedures (including what annotation to modify when superseding an item, how to manage user permissions and editing items after approval) are documented on the wiki too:

<https://www.wiki.ed.ac.uk/display/datashare/Procedures>

- DSpace Simple Archive Format

<https://wiki.lyrasis.org/display/DSDOC6x/Importing+and+Exporting+Items+via+Simple+Archive+Format>

- DSpace export to AIP

<https://wiki.lyrasis.org/display/DSDOC6x/AIP+Backup+and+Restore>

Reviewer Entry

Reviewer 1

Comments:

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
Accept.

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

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 DataShare Preservation policy sets out the responsibilities of the repository including long-term preservation. DataShare undertakes to retain archived data indefinitely. We do not take deposits with defined retention periods.

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/preservation-policy>

This policy indicates that “Items will be retained indefinitely”, and that the repository will try to ensure continued readability and accessibility of items (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/preservation-policy>). Repository policies were initially created using the OpenDOAR policy tool (<https://v2.sherpa.ac.uk/opendoar/policytool/>) and have been reviewed each time we have applied for trusted repository certification.

While the policy does not commit us to operating the DataShare service forever, it does commit us to attempt to migrate the content to an appropriate repository if the service needed to shut down in future (exit strategy). DataShare staff contributed to two previous successful whole-repository migrations, one from EDINA’s OpenDepot service to an external public repository (Zenodo), and one from EDINA’s ShareGeo repository to Edinburgh DataShare (<http://datablog.is.ed.ac.uk/2017/04/18/edinas-sharegeo-open-content-into-datashare/>).

Depositors are encouraged to ‘future-proof’ their work and choose sustainable file formats, as in the Checklist for Deposit (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>). We receive deposits in a number of formats as a multi-disciplinary repository, and levels of support for types of file formats are documented on our web pages as “Choose the best file formats” (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats>). Users only rarely make deposits of files in bulk, so

we don't have a fully automated process. Curators ensure file formats are accessible and sustainable where practically feasible (preferably open source or international standard formats) ie in the vast majority of deposits:

<https://www.wiki.ed.ac.uk/display/datashare/Procedure%3A+New+File+Formats+in+DataShare+Submissions>

The Depositor Agreement sets out the responsibilities of the submitter and grants permission to the repository to take such steps as may be necessary for the preservation of the data, metadata and documentation:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/depositor-agreement>

DataShare is doing high-level preservation. The approach is risk-based. Our Checklist for Depositors sets out how depositors should evaluate the significant properties of their data and address them, by structuring the data, and producing documentation to aid navigation and access, and selecting sustainable file formats appropriate to what their data does, to ensure long-term accessibility and sustainability of their files so they will choose a format that preserves the significant properties. In addition our curators contribute to the sustainability of the files by meticulously checking and documenting file formats in new submissions, in case our users have not had the time to follow our advice. See our file formats registry and procedure for new file formats below. We gather information about software dependencies such as whether the files have been compressed. The team discusses and reaches a joint decision on whether new formats are classed as 'supported' ie recommended (the more sustainable formats), merely 'known' ie acceptable or whether we advise they be converted to a more sustainable format. We also consult Library of Congress, PRONOM and so on, taking into account whether the format is a community standard and whether it is open source, whether it is plain text, in other words the significant properties. This collected information is recorded in DataShare's integrated file format registry (part of the DSpace system) and the wiki and used to keep our page on choosing the best file formats up-to-date. On that page we advise users on which formats should be avoided and which are best converted. Thus we do not rely merely on the extension to identify the file. In fact our registry allows us to distinguish more than one format with identical extensions, for example '.dat' and '.raw' are documented as matching more than one format in the wiki copy of the file format registry.

We also manually open the files of every deposit to ensure it can be opened and used, in effect this is another way that we check the significant properties and carry out our own validation. In a small minority of cases we won't open the files if expensive proprietary software unavailable to use would be needed. In the case of submissions containing hundreds of files, we will open a representative sample of files, since it would not be proportionate to the level of risk to spend the time to open every single one. In our risk-based approach, it makes sense to rely to a certain extent on the researcher to deposit a file that works and is high quality. But in future we plan to implement a digital preservation solution that automates most of these checks.

The RDS team has agreed with colleagues from across the Library and University Collections division on a plan both for short-term and longer-term steps to protect the integrity of our data. The plan is recorded in our wiki:

Digital preservation and DataShare

<https://www.wiki.ed.ac.uk/display/datashare/Digital+Preservation+and+DataShare>

Short-term: migration of priority file formats

We have identified a list of file formats we consider at greater risk of becoming inaccessible, and which should be migrated to new versions or more sustainable formats. These were identified based on expertise within the RDS team, including the results of the investigations we carry out when evaluating novel file formats submitted to the repository, as documented in our file format registry. We have also drawn on resources such as the Library of Congress Recommended Formats and PRONOM. In 2020 we identified and compiled a list of two hundred and fifty files requiring migration including RAR archives and old Microsoft Office formats. The Research Data Support team has manually migrated these files to the new versions of the formats, or in some cases new formats (using MS Office and 7-zip), and added the bitstreams manually to the Items in DSpace. In all cases, we stored the details of the curator's actions, and the choice of new format, in a publicly accessible text file added to the Item (a minimal equivalent, analogous to the METS file that automated systems would produce). The team informally describes this process as 'artisanal digital preservation'.

We do not currently deal with AIPs in bulk, but the files and metadata we store are equivalent to AIPs. Only a handful of data (seventeen datasets) are more than ten years old, so the vast majority of our datasets (approximately two thousand nine hundred) are not old enough for the risk to be significant enough for the data to be done in bulk. We have a great diversity of content, so migration in bulk is not appropriate.

EXAMPLE: dataset where one file has been migrated, and the new file added alongside the original, and the preservation action has been documented:

Robertson, Neil. (2016). High Absorption Coefficient Cyclopentadithiophene Donor-free Dyes For Liquid and Solid-state dye-sensitized Solar Cells, [dataset]. University of Edinburgh. School of Chemistry.

<https://doi.org/10.7488/ds/1410>

EXAMPLE: Deposit where many files were migrated, all documented also in the preservation action text file as part of the deposit 'DataShare_curator_preservation_action.txt':

Gibson, Lorna; Paul, Laura; Sudlow, Cathie LM. (2017). Potentially serious incidental findings on brain and body magnetic resonance imaging conducted among apparently healthy adults: a systematic review and meta-analysis, [dataset].

<https://doi.org/10.7488/ds/2100>

EXAMPLE: Deposit originally containing seven files, just one was in a format identified on our list as needing migrated, and this was added:

Remijisen, Bert; Impey, Angela; Ajuet Deng, Elizabeth Achol; Deng Yak, Simon Yak; Ayuel Ring, Peter Malek; Penn de Ngong, John; Reid, Tatiana; Ladd, D. Robert; Meyerhoff, Miriam. (2012). Manyang Deng, 2009-2012 [sound]. University of Edinburgh. School of Philosophy, Psychology and Language Sciences.

<https://doi.org/10.7488/ds/134>

PRESERVATION PLANNING - Longer-term plan: more robust file format identification, and potential automation

For the preservation of our data beyond 2021, and into the long-term, we plan to do the following:

- Carry out at least a biennial survey of our data evaluated against both our own extensive DataShare file format registry

and the PRONOM file format database. We will use an automated system such as DROID or Archivematica to produce a report to identify any Items in DSpace containing file formats that we decide should be migrated.

- We will work with colleagues from the University's Digital Library team and Centre for Research Collections who have considerable expertise in DROID and Archivematica. Colleagues are currently working on integrating Archivematica into workflows for other systems run by other teams. Integration with DataShare would be consistent with the license held for the software. We will build on that work. We aim to integrate Archivematica with DataShare, so that we may automate both the identification of file formats (not just at Item level in DSpace, but also files contained within compression archives such as zip) and the normalisation of files.

Library and University Collections is currently developing an overarching digital preservation policy, led by the recently appointed Digital Archivist. A harmonised policy on recommended file formats is in development within a Digital Preservation Working Group which members of our team both chair and are represented upon. When this policy is approved, we will ensure DataShare submissions are checked and manually normalised accordingly. We will also from that point forward apply the policy in our subsequent biennial survey and migration work.

Checklist for deposit [advice to depositors]

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>

Choose the best file formats [advice to depositors]

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats>

Checklist: Checking a new Item Submission to DataShare - datashare - Wiki Service [See point 3 " Are the file formats Supported?"]

<https://www.wiki.ed.ac.uk/display/datashare/Checklist%3A+Checking+a+new+Item+Submission+to+DataShare>

A separate procedure is in place if the evaluation of the submission discovers a file format not previously archived with us:

Procedure: New File Formats in DataShare Submissions - datashare - Wiki Service

<https://www.wiki.ed.ac.uk/display/datashare/Procedure%3A+New+File+Formats+in+DataShare+Submissions>

Wiki page: File Format Registry (to allow us to search for filename extensions, we keep a copy of the content of our DSpace file format registry, manually compiled, on our wiki):

<https://www.wiki.ed.ac.uk/display/datashare/File+format+registry>

Reviewer Entry

Reviewer 1

Comments:

The new evidence makes the process much clearer and proves a structured approach on preservation planning. Planned activities to improve and automate the process and use Archivematica as a digital preservation system are convincing. You may consider file format validation (if applicable) in your longer-term plan, which isn't mentioned there yet.

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
Accept

Reviewer 2

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

Response:

Approach – The Research Data Support team applies standards to data and metadata quality commensurate with a FAIR-conformant repository. We encourage and facilitate the inclusion of high quality and thorough metadata and documentation with deposits. We aim to obtain sufficient metadata and documentation to permit reuse of the data by peers unfamiliar with the data. At a minimum, we require at least one human-readable documentation file to explain the dataset/data item. (See our definition of dataset, <https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/definitions>). Quality control – The DataShare submission form automatically ensures that submitters provide at least a minimal set of mandatory metadata describing the data, and optional fields allow for a rich set of metadata values to be captured. The PDF document, Edinburgh DataShare User's Guide, shows the fields that are required and optional:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository> (link from bottom of page).

Every dataset submitted to DataShare is checked by a curator from the Research Data Support team to check the documentation and metadata provide enough information at least to allow another researcher from the same field to re-use the data. As a general rule this means deposits should have a dataset description providing an overview of the data, if there is not yet an associated publication, and usually a readme.txt file containing information about the files and

their contents, and how they relate to each other. Curators encourage submitters to aim to provide sufficient documentation for potential users from beyond their own research community, time permitting. The curation team has a high level of expertise in a variety of research fields, as well as in-depth knowledge of research data management, allowing us to make an informed evaluation of the quality of metadata and documentation on the dataset submissions we receive from the very wide variety of disciplines practised at the University of Edinburgh. Only items with accompanying documentation are approved for ingest into the repository. On the frequent occasions where the initial submission falls short of the optimal level of metadata or documentation, or where some information appears to be inconsistent or potentially confusing, the curator will enter into a dialogue with the submitter, through our service desk call management system, to encourage them to provide rich metadata and documentation. In the rare event the submitter is unresponsive, curators will apply the minimum standards, but without sufficient information for re-use, we will reject the submission. In the vast majority of cases, submitters are happy to provide further documentation, and their deposit is then approved.

Citations of associated publications are treated as an important component of the metadata. In the many cases where the data support a manuscript that is still in submission, a note to this effect is included. Where a journal publication occurs after approval, and if the depositor fails to send us the details, curators often make use of the University's CRIS and the associated web portal, to find and link journal publications to our datasets. When a valid URI is included in the `IsReferencedBy` field, DataShare has a special feature that adds a hyperlink to the deposit landing page (summary view) for the convenience of users. We sometimes tweet our congratulations and links when a depositor's manuscript is finally published.

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 the team's 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. New members of the team undergo detailed training and practice exercises to make sure they know and understand the standards to be applied.

User feedback – Should a user have difficulty reusing data, for example because of a missing file, the contact details of the Research Data Support team are available through the Contact button.

The DataShare homepage has a prominent link to this page "How to deposit" which tells submitters to include documentation and that they should complete as many metadata fields as possible:

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/how-to-deposit>

Our "Checklist for deposit" sets out how users should meet our standards by labelling files clearly (Step 3 – Prepare your data) and by providing documentation (Step 4 – Prepare your documentation):

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>

Our curators' checklist sets out how we check data and metadata quality including point 5 (Are the files corrupted?), point 6 (Is the metadata sufficient? Lists six points to look at), point 10 (Is there enough documentation?):

"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>

- Metadata Schema

<https://www.wiki.ed.ac.uk/display/datashare/Current+metadata+schema>

Reviewer Entry

Reviewer 1

Comments:

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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The Research Data Service provides training on research data management across the university, through which we help users to see their research data management activities as part of a research data lifecycle. Archiving and sharing data is presented as an integral part of that lifecycle, and the DataShare open access data repository is promoted as a supported tool allowing our researchers to achieve good practice in RDM through FAIR sharing.

Accordingly, the home page of the Research Data Service guides users to appropriate services through a lifecycle approach: “Before you begin”, “Research in progress”, “Approaching completion”

(<http://www.ed.ac.uk/is/research-data-service>).

DataShare's workflow is encapsulated on the repository homepage:

“Edinburgh University researchers who have produced research data associated with an existing or forthcoming publication, or which has potential use for other researchers, are invited to upload their dataset for sharing and safekeeping. A persistent identifier and suggested citation will be provided.” <https://datashare.is.ed.ac.uk>

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. The workflow is described in our web pages, which make clear that submitters may provide their data, metadata and documentation to the repository, for ingest, sharing and preservation. See below 'Checklist for deposit' which guides the depositor through steps they should carry out before submission to ensure the structure, format and so on of their deposit are optimal for accessibility and reusability.

When data are submitted to the repository, the preservation workflow starts. The first step is the ingest of data assets. Depositors are prompted to follow the steps in our 'How to deposit' page, with links to further information in the 'Checklist for deposit' and the 'Edinburgh DataShare User Guide'. Moderation of the deposit follows a workflow of its own, implemented by the curators. See below the link to this procedure on the DataShare wiki "Checking a new item submission". All procedures are documented on the DataShare wiki, and these include creating and moving collections, adding users to groups, editing items following a user's submission, bulk metadata editing (for large deposits submitted in batch mode), superseding an item, and withdrawing an item (tombstoning a record). In some cases a dialogue is opened with the depositor to suggest and agree appropriate changes to a deposited item before accepting into the repository.

Depositors receive a standard automated email when their deposit is ingested, confirming that their submission has been approved and archived, and providing them with a reminder of the Depositor Agreement and a link to it, and with contact details for the Research Data Support team. An automated email is also sent to remind the depositor when their embargo is about to expire. The text of these messages is kept in the team's GitLab repository so that any changes to these depositor communications are managed by the team and documented.

Once ingestion of data is final, the dataset is assigned a persistent identifier with an accompanying optional licence and a suggested citation. DSpace is optimised for discovery by search engines and Google Scholar and is OAI-PMH compliant allowing harvesting by aggregation services including IRUS-UK and Clarivate's Web of Science Data Citation Index.

In terms of archival data transformation, we are carrying out short and long-term work to systematically ensure the longer-term accessibility of our data. Please see response 10 for our current and future activities on this point.

The University's RDM Roadmap process, our Steering Group and our Communications Plan (internal document) all provide for consultation with appropriate stakeholders for change management around any of these workflows.

- How to deposit

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/how-to-deposit>

- Checklist for deposit (aimed at depositors)

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/checklist>

- Submission policy

<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/service-policies/submission-policy>

- Checklist: Checking a new Item Submission to DataShare (aimed at curators)

<https://www.wiki.ed.ac.uk/display/datashare/Checklist%3A+Checking+a+new+Item+Submission+to+DataShare>

- DataShare Procedures

<https://www.wiki.ed.ac.uk/display/datashare/Procedures>

- Automated emails

<https://www.wiki.ed.ac.uk/display/datashare/Automated+Emails>

- Digital preservation and DataShare

<https://www.wiki.ed.ac.uk/display/datashare/Digital+Preservation+and+DataShare>

Reviewer Entry

Reviewer 1

Comments:

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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository makes the data highly discoverable, contributing to the robustness and integrity of our scholarly record and to the 'FAIRness' of the data item, through the following means:

- The metadata schema includes discovery metadata including keywords, subject classification, spatial and temporal coverage.
- The metadata schema is based on the Dublin Core Metadata Initiative (DCMI) international standard and meets the minimum requirements of the DataCite international standard.
- The metadata may be searched using DataShare's search box.
- Faceted browsing is available and tailored to every page of the repository, allowing end-users to discover subsets of data by browsing Data Creator, keywords, spatial coverage, funder, year of accession and data type eg 'image'.
- The repository complies with the OAI-PMH protocol to enable the indexing and harvesting of the metadata and the online dissemination of data records. The repository is tuned to be highly discoverable in both Google Scholar and Google Dataset Search. Technical steps have been taken to remove all possible barriers (known to us) to allowing DataShare deposits to become findable in Google Dataset Search. Namely, the technical support team has tagged all our Items with the relevant schema.org so that Google may harvest the Items. New deposits will also be automatically tagged in the same way.
- As a result of our DOI minting, all our datasets are listed on DataCite Search:
<https://search.datacite.org>
- Curators support depositors to create richer metadata.
- Curators transfer metadata to the University's CRIS through a semi-automated process, to link the data to profiles of the data creators, projects and papers, to enable discovery and integration.

The repository mints a DataCite DOI persistent identifier for each data item or dataset, as well as providing handle persistent identifiers for each dataset and collection. The repository provides an example data citation on the landing page (summary view) of each dataset. This data citation is generated automatically by the repository and includes the data creators' names, dataset title, year of accession and the DOI persistent identifier.

The repository is listed in the Directory of Open Access Repositories (OpenDOAR), the Registry of Open Access Repositories (ROAR) and in the Registry of Research Data Repositories (re3data).

- Metadata schema

<https://www.wiki.ed.ac.uk/display/datashare/Current+metadata+schema>

- The metadata is added to the University of Edinburgh's CRIS, the contents of which are publicly available online (search

for 'datashare'):

<https://www.research.ed.ac.uk>

- DataCite search – datasets with DOIs from the University of Edinburgh (10.7488)

<https://search.datacite.org/repositories/bl.ed>

- Edinburgh DataShare listing on re3data

<https://www.re3data.org/repository/r3d100000047>

- Edinburgh DataShare listing on OpenDOAR

<http://v2.sherpa.ac.uk/id/repository/1176>

- Edinburgh DataShare listing on ROAR

<http://roar.eprints.org/3007>

Reviewer Entry

Reviewer 1

Comments:

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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

Overview

Edinburgh DataShare is a discipline-agnostic data repository. Depositors are encouraged to make their data FAIR through inclusion of adequate documentation and by selecting appropriate file formats that maximise the long-term accessibility and sustainability of their data. We communicate with depositors through information provided on our web pages, training, and through dialogue arising directly from the curation workflow. Our Checklist for Depositors sets out how depositors should evaluate the significant properties of their data and address them, by structuring the data, and producing documentation to aid navigation and access, and selecting sustainable file formats appropriate to what their data does, and the standards agreed by their research community, to ensure long-term accessibility and sustainability of their files so they will choose a format that preserves the significant properties.

Curator actions to enable reuse of the data over time

Each individual deposit undergoes hands-on curation by a trained member of the Research Data Support team. We check the dataset is understandable and interoperable for different operating systems etc.

We maintain a detailed file format registry. While our DSpace repository categorises formats as 'Supported', 'Known' or 'Unknown', we have found it more useful for the purposes of communicating with users to classify formats as: those we recommend; those we consider acceptable; and those we ask users to convert to something more sustainable. The research and evaluation process by which we arrive at that classification is documented on our wiki:

Procedure: New file formats

<https://www.wiki.ed.ac.uk/display/datashare/Procedure%3A+New+File+Formats+in+DataShare+Submissions>

Curators have access to detailed notes on our wiki to help them:

- Identify novel formats;
- Work together as a team to evaluate formats, using external sources of expertise (PRONOM, Library of Congress and Wikipedia) and settle on how to advise users about them.

To address the reusability of the data over the long term, taking account of the possible evolution of file formats, we are taking steps to assess the existing data periodically, and take preservation actions where appropriate. For example, see the short term and longer term recent and planned activities around file format migration in Response 10, Preservation Planning.

To help curators identify novel formats, we record all our notes on formats, filename extensions and recommendations on our wiki, so it is easily searchable. This page duplicates the information in our DSpace file format registry, which is not searchable: <https://www.wiki.ed.ac.uk/display/datashare/File+format+registry>

Steps we ask depositors to take to enable the reuse of the data over time

Advice to depositors on our web pages: "Choose the best file formats" (the Checklist for deposit directs users to this page) <https://www.ed.ac.uk/information-services/research-support/research-data-service/after/data-repository/choosing-file-formats>

The Checklist for Deposit has been designed to advise new depositors and those researchers interested in proactively making their data more usable and reproducible. As the depositor is the best expert on the data item, we ask them to consider a number of questions before they make the step of uploading their data, which will enhance the usability of the

submission. These include:

- "Are you depositing multiple datasets or items? (Setting up a new collection for a research group or research project may be appropriate.)
- "Have you structured and labelled your data in a consistent manner?"
- Which format(s) will you deposit to help ensure the broadest accessibility by others now and in the future?
- Are there any discipline specific data formats? Does Edinburgh DataShare specifically support those format(s)?
- Have your data been created, edited or compressed with specific software(s)?
- Have you prepared documentation files for your dataset(s)?
- Have you spelled out acronyms and explained the labels of your variables and values?
- Have you included research methodology reports and any other relevant information?
- Do you have data citations ready for any underlying "source" data (such as base maps)?
- Have you sufficiently anonymised your data, or obtained explicit consent from any data subjects whose identity could be revealed from the data (including images)?
- Does your data item require an embargo period before it is made open access?
- Have you considered the use of an open licence for your data?
- Do you know about the default open licence used for repository material?
- Are you prepared to fill in a rights statement, if the "No licence" option is chosen?

Explanatory note regarding the latter questions about open licences: The repository adds a CC-BY 4.0 International Licence by default, with an option to override this choice during the deposit workflow. Curators check the licence on submissions, and strongly encourage all depositors to choose an open licence. We have found this a much more successful way to achieve FAIR-conformance in the items in the repository than to leave it up to the depositor to specify a licence from scratch, since some may not be aware of the value of applying open licences.

Reviewer Entry

Reviewer 1

Comments:

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:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

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

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The repository uses the popular open source DSpace software, which has a thriving community of users and developers (over 1,000 instances worldwide). The software runs on Linux servers (virtual machines) maintained by the University of Edinburgh. DataShare's developers and Service Owner maintain close contact with the infrastructure team to ensure high levels of capacity and availability. We have begun measuring overall availability through an uptime checking tool which 'pings' the live system every ten minutes. A network procurement has been agreed, bringing faster and more reliable network speeds across university spaces.

As part of the Information Services Group of the University, these teams apply the ITIL framework of service management best practice in the delivery of IT services. All upgrades to the software, whether developed in-house or provided by the DSpace community, are managed through the ITIL processes (incident management, change management, release management, problem management). In order to maintain availability of the service for users, releases are assessed in terms of risk of causing disruption to users and other services, and are timed to minimise potential disruption to users. Development on the system is managed through an issue tracking system and fortnightly team meetings, with releases planned ahead to implement priorities identified in the Research Data Management Roadmap, as well as fixing bugs or enhancing usability in response to user feedback.

Information Services Group (the organisational unit in which all repository staff are employed)) has a business continuity plan in place. During the Covid-19 social distancing measures, our offices have been closed, but as the business continuity plan has been put in place, has all of the staff have continued to develop the system and to support users, to curate new deposits and maintain existing metadata. Essential infrastructure staff have had continued access to the data centres where the DataShare servers reside, such that there has been no interruption of service to depositors and no interruption of availability of the data and metadata.

Over the lifetime of the repository, since 2008, our depositors have needed to share ever larger datasets in greater numbers, as lab equipment and video cameras among others generated ever greater volumes of data. We have enhanced the DSpace system to allow deposit and download of larger file sets, now up to a maximum of 100 GB per

dataset. This required us to add significant new features ('drag and drop' upload to allow larger file sets to be deposited through the web interface, and zipped file sets with a special download-all button to facilitate downloads via the browser, bypassing the database to reduce the risk of downloads being interrupted). As described on our blog, deposits above 20 GB need to be imported through our batch import process – the team maintains an increasingly large staging area and uses various transfer routes for the files to keep on facilitating the larger deposits coming to us through this route.

- Extensive technical documentation of Edinburgh DataShare's codebase, infrastructure and customisations:

<https://www.wiki.ed.ac.uk/display/datashare/dev>

- Lyrasis makes detailed documentation of DSpace available:

<https://duraspace.org/dspace/resources/>

- Business continuity in practice: through the blog the Research Data Support team encouraged users to continue to deposit data during the Covid-19 lockdown in Scotland:

<http://datablog.is.ed.ac.uk/2020/04/16/research-data-management-in-a-time-of-quarantine/>

- "DataShare 3.0: The 'Download Release' means deposits up to 100 GB"

<http://datablog.is.ed.ac.uk/2017/10/19/datashare-3-0-the-download-release-means-deposits-up-to-100-gb/>

Reviewer Entry

Reviewer 1

Comments:

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

Accept

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

The University has a Chief Information Security Officer and Information Security team available for consultation at any time. This team also monitors risks to the university's cyberinfrastructure continuously, provides warnings to users about, e.g. phishing attacks, provides informational web pages and regular training courses to the university community (<https://www.ed.ac.uk/infosec>). All of the repository staff have taken their online course, Information Security Essentials, and are encouraged to take further information security training as part of regular professional development.

The development, delivery and continual improvement of the DataShare service are managed within the ITIL service management framework, including the assessment of risks to the data and the service each time a new change is released.

Our experience of running DataShare as an open data repository since 2008 shows that the threat of malicious access is relatively low. Nonetheless, the University login is used to restrict access for submitting data, and to restrict admin access to the curators and the technical team. To create an account on DataShare, a user must have a University of Edinburgh login. This is a two-factor authorised system based on the UK Access Management Federation for Education and Research, run by Jisc (<https://www.ukfederation.org.uk/>).

The Edinburgh DataShare repository contents are backed up to tape daily and database snapshots are available on disk for two days. Database snapshots are also backed up to tape. Daily tape backups are retained for 30 days and a monthly archive copy is kept for a year.

The repository incorporates a virus checker which checks all submissions that come via the web form. The repository has been moved from HTTP to HTTPS for added security of all users.

We store an extra copy of each file external to DSpace in order to allow us to provide a download feature that has capacity to allow downloads of larger files (up to 100 GB). Users access this via the 'Download all' button.

Reviewer Entry

Reviewer 1

Comments:

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:

- Strong community engagement.
- The publicly available information about internal processes on the wiki and your file format registry is really appreciated and strengthens credibility of the documentation.
- The amount of information provided in several media types for producer guidance is really impressive and can't be found on every repository.
- Good f

Reviewer 2

Comments:

It is an excellent application.