



University of Edinburgh Research Data Service Roadmap, 2017-2020

Report on Objectives and Achievements/Gaps

Priority settings as designated in 2017: colour coded in Red, Amber, Green to note degree of achievement with description of activities/milestones/deliverables by category. Achievements/gaps discussed at the Research Data Service Steering Group, 1 October, 2020, to inform next Roadmap.

A. Unification of the Service

1) Devise a unified look and feel to the service components, as part of Digital Research Services.

Priority: High

- Research Data Service website maps service components to simplified lifecycle, without undue jargon/brand names. User tasks were tested with consultation from University web team in 2018 to audience of researchers new to the University; improvements edited into web pages. All pages refreshed by team, summer 2020.
- Digital Research Services website built with 'deep links' into research data service components and simplified journey from the home page (published 2019).
- Service brochure revised, printed June 2019, used in face to face training before COVID-19.
- Professional short promotional service video published, Sept. 2017.
- Case study interviews with university researchers published Sept. 2018 (innovation fund deliverable with student intern).
- Quick Guide series introduced 2019 to provide concise, readable guidance on specific topics.

2) Create a coherency to service as a whole and its underlying components.

Priority: High

As above.

3) Offer a simplified user journey starting from the home page.

Priority: High

As above.

4) Investigate and deliver on options to allow interoperability between systems, e.g. to transfer data easily from one system to another.

Priority: Medium

- Provision of "Tier 2" storage for biomed research activities to allow longer term/higher capacity close to compute which is required for this type of workload.
- DataShare interoperability improved with automatic ingest of records into Pure plus harvesting in Google Dataset Search, Google Scholar, DataCite, Altmetrics, and Web of Knowledge Data Citation Index. Altmetrics and user statistics available per item; proof of concept achieved with ORCID.
- DataVault interoperability with DataStore and Pure.
- RSpace interoperability with DataStore, DataShare, and protocols.io.
- Data Safe Haven interoperability explored with DataVault; further technical and information security discussions needed.



- Light touch benchmarking occasionally reported as part of engagement with peer institutions/conferences.

5) Horizon scan on options for replacement/merging of system components.

Priority: Low

- Light touch reviews were done for DataShare and DataVault (2017).
- Professional engagement ensures team's knowledge of repository systems capabilities
- Review of large scale distributed erasure encoded storage systems as possible technology replacements for DataStore.

Known gaps for Unification of Service:

- Full service SLA not yet completed; still working on service levels of component parts.
- Still some confusion by users regarding different systems under the service umbrella due to naming convention with use of prefix 'Data'.
- Reasonable interoperability between systems now but not to level of 'end to end service'. Could learn more about user requirements for this.
- Full horizon scan not completed; will attempt synergy with planned procurement for new CRIS (Pure or replacement) after the REF.
- Research Data Management Policy will be ten years old next year; a rewrite or refresh will be an opportunity to address some unification aspects of the service.

B. Data Management Planning

6) Increase the number of fully-costed plans written by researchers across schools, including University-funded research.

Priority: High

- The number of plans being written in DMPonline has increased, reaching 393 in 2019.
- The RDS team is increasingly being asked to review plans, as measured in Unidesk statistics.
- The library of UoE plans is now available from within DMPonline and continues to grow.
- The guidance and support content in DMPonline is being reviewed and updated annually.
- The UoE default templates have been fully updated to reflect changes in both policy and practice within the University.
- Through DRS funding we have subscribed to the enterprise version of the tool for full advantage to the institution.
- All requests for support with a DMP have been responded to quickly and most are resolved to the user's satisfaction within a week. Some complex plans take longer due to meetings with researchers and multiple iterations of the plan, but again satisfaction is high.

7) Increase the number of schools/departments that require a simple DMP as part of a PhD research proposal (from zero).

Priority: Medium

- The team provides tailored DMP and RDM training for a number of schools, including Geosciences, Business, SPS, IGMM and Education, and efforts are ongoing to reach other schools.



- Expectation introduced in one or two schools (GeoSciences, Business) of a DMP to be part of a PhD research proposal, if not a hard requirement.
- Work undertaken with Roslin Institute to ensure archiving destinations for PhD students' work as part of thesis deposit.

8) Encourage software management planning as part of DMP.

- Actively collaborating with SSI and EdCarp to deliver data/software carpentry training to colleges.
- MANTRA and scheduled training courses have been revised to include the importance of sustainable software practices.
- This topic is being addressed through Library Research Support's engagement with the University's Open Research agenda.

Priority: Medium

9) Explore and evaluate possibilities for 'active' data management plans.

- Awareness of wider developments through professional development/conferences.

Priority: Low

Known gaps for Data Management Planning

- The team does not have oversight of all DMPs submitted with grant proposals to monitor inclusion of costings, rather we tend to see only ones where our assistance is sought. We are still trying to establish processes for the timely sharing of information between ourselves and ERO about DMPs, as well as utilise information from Worktribe effectively.
- A Data Protection Impact Assessment tool has been introduced as part of the Ethics workflow/toolset but production of a DMP remains outside of this process.
- As yet we have not had any schools move forward with customising DMPonline, though we continue to offer this service; most schools seem satisfied that the guidance we provide is sufficient. Nor have we had any schools interested in writing SOPs for data management.
- No school has yet decided to provide a bespoke DMP template for their PGR's; the large majority of schools still do not mandate a DMP with research proposals.
- DMP templates do not cover caring for code over the course of the lifecycle to the same extent as data, though this is typically the case. There is we are working on this through LRS's engagement with the Open Research agenda.
-
- No automation, or 'triggers' have been established in terms of active data management plans or interventions during the research project lifecycle.

C. Working with Active Data

10) Provide secure setting for sensitive data and set up controls that meet ISO 27001 compliance and user needs.

- Facility established as part of Research Data Service, team structure involving both RSS and RDS teams, with soft launch at Dealing with Data event, November 2018.
- Technical build with innovative security approaches part of major DRS-funded development project over the Roadmap period. Penetration tested, continuing on annual basis.



- ISO 27001 standard compliance established with robust controls in place, a full technical and operational documentation set, and dedicated roles. The DSH has been independently audited as an Information Security Management System by Lloyd's Register and received ISO 27001 certification in December 2019, with ongoing internal and external audits continuing.
- The facility has been approved by NHS Digital and approval expected shortly with NHS Lothian. Other external data owners such as National Register of Scotland, Police Scotland and AstraZenica have also given approvals.
- A small number of research projects have been onboarded and are using the facility.
- A short-term remote access solution was put in place for essential research following COVID-19 working from home, with a full solution prioritised and to be rolled out by end of 2020.

Priority: High

11) Modernise the code versioning service component.

Priority: High

- Gitlab has been introduced; received by users with enthusiasm.
- User requirements for Subversion have been sought with a view to retiring the system.

12) Ensure usable interface for researchers and their collaborators to share storage space for research data.

Priority: Medium

- Datasync has been upgraded; longstanding bug has been fixed, documentation updated.
- The team has added information about Confluence Wiki and Sharepoint to collaboration web page and in training, to suit different types of research projects.
- A quick guide about storage options was published on the service website and was widely circulated during COVID working from home.

13) Advocate and understand how lab-based research settings can move to electronic lab notebooks.

Priority: Medium

- RSpace trial completed and moved to a brokered licencing arrangement with user support SLA with ResearchSpace. Work done with Procurement Office, Information Security team, Software Services, as well as the Digital Research Facilitators and IS college liaisons to understand requirements.
- Steering group approved a plan for evaluating external research tools with suitable criteria for adoption.
- RSpace user statistics now incorporated into regular metrics, with steady growth.
- Webinars have been held for each college to show how the tool can be used in research environments. Electronic notebooks included in RDM training materials.
- A new trial for protocols.io was undertaken in December 2019, with similar characteristics.
- An open research tools web page draws attention to the two external tools as well as EDINA's Noteable and Biosciences' Wikibench.

14) Ensure researchers doing database activity have secure, networked backed-up resource.

Priority: Medium



15) Provide coherent, safe solution for University researchers who need to encrypt their data in Research Data Service systems.

Priority: Medium

- Team members have been providing bespoke advice to research teams regarding safeguards such as encryption for personal and sensitive data, liaising with Information Security team and the DPO as needed. (The DPO has been sending enquiries to the team as well.)
- Fuse environment supported on EDDIE storage.

16) Ensure the stability of the University shared storage space and plan for growth.

Priority: Low/Business as Usual

- Infrastructure refreshment has taken place and cost recovery models have worked well.
- In some cases institutes have organised hardware purchases through the central service (eg. IGMM renewal approximately £950,000).

17) Assist researchers who require datasets for analysis from all disciplines.

Priority: Low/Business as usual

- The former Data Library catalogue was reduced and streamlined to a short set of web pages including links to data search engines and a few well-known data sources, “Discover and re-use data” (<https://www.ed.ac.uk/information-services/research-support/research-data-service/during/discover-and-re-use-data>).
- Currently pursuing potential collaboration with EFI, EIDF and DCC to understand data catalogue requirements.

Known gaps for Working with Active Data

- Dissatisfaction from some parts of the research community with existing costs/cost recovery model.
- Researchers working with external partners, particularly with sensitive data or desiring to conduct open research, require support for understanding potential platforms and tools.
- ‘Database as a platform’ was not pursued as it was not adopted by the DRS programme. Schools may need to provide local IT support for this.
- Regarding encryption there is University advice but not a key management service. This is out of scope for the Research Data Service to provide.

D. Data Stewardship

18) Provide enhanced display and preview of multimedia and numeric files held in repository.

Priority: High

- Work on this is scheduled but has been delayed due to the need to migrate from EDINA to L&UC infrastructure and support.

19) Alternative to open sharing exists with immutable, replicated secure long-term storage.

Priority: High



- DataVault facility was successfully launched in January, 2019 following stages of DRS-funded project development to take the Jisc-funded prototype to suitable levels of functionality, including encryption, metadata capture and roles/permissions.
- User documentation has been written along with videos and training for support staff.
- A review process and 'dashboard' was agreed and implemented to ensure university management of data when the original data owner leaves the institution.
- In Spring 2020 data migration from the 'interim DataVault' was completed.
- In August, 2020 performance and speed were enhanced to allow up to 10 TB deposits.

20) Communicate options to researchers for data sharing and curation customised to their situation (discipline, funder, project longevity).

Priority: Medium

- This is done directly through helpdesk enquiries and in training sessions as well as providing general advice through the website and online learning materials.

21) University repository that provides FAIR (findable, accessible, interoperable, reusable) principles in line with emerging funder and international principles (EU Open Science Cloud etc.).

Priority: Medium

- Participation in COAR and RDA working groups to keep up with international standards and expectations on FAIR. A number of conference presentations and papers to highlight UoE approach at international conferences, both peer reviewed and by invitation.
- As part of Core Trust Seal application for DataShare (see below) alignment with FAIR was demonstrated.
- A virtual collection of COVID-19 related datasets was brought together in 2020 following FAIR and open principles in DataShare: <https://datashare.is.ed.ac.uk/handle/10283/3749>

22) Repository makes best use of metadata standards and citations to improve impact for researchers.

- The team continues to seek the best metadata possible and linkage with articles when moderating deposits: a total of 735 items were submitted and reviewed according to our quality control procedures during 2017-20 so far.
- Altmetrics badges added to DataShare items with one or more citations.
- Users encouraged to create metadata records in Pure for datasets not lodged with a university service.
- Metadata schema updated to meet DataCite 3.2 metadata schema.

Priority: Medium

23) Embrace new technologies and metadata standards for curating and providing access to research images.

Priority: Low

- Omero as a service being explored by RSS with a school potentially providing support.

24) Repository continues to meet trusted digital repository requirements for researchers and actively migrates supported file formats for long-term preservation.

Priority: Low / Business as Usual



- Application for trustworthy digital repository certification from Core Trust Seal submitted in July, 2020, awaiting outcome from peer reviewers. This will update previous (2014) certification from Data Seal of Approval and align with FAIR principles.
- The team migrated known obsolete file formats in DataShare for the first time in June, 2020.
- Collaboration with Digital Library explored to streamline future file format migrations using Archivematica.
- Work scheduled to deposit a complete snapshot of DataShare contents in the DataVault as a disaster recovery option.

Known gaps for Data Stewardship

- Some prioritised work on DataShare (data and image previews, APIs and Linked Open Data for data scientists, metadata schema overhauls), has been delayed due to bug fixing and the need to migrate from EDINA infrastructure and support to Digital Library as well as prior delays in the DSpace release schedule. When migration is complete (release scheduled for October) the domain will change to <https://datashare.ed.ac.uk> (from datashare.is.ed.ac.uk) as requested by the steering group.
- DataVault has filled a gap for both restricted access data and larger datasets. More infrastructure and support for users wishing to share both of these may be needed e.g. SFTP or GridFTP with disc space to host these; information governance support such as writing data sharing agreements, automating processing of requests for access.
- Outreach and guidance to researchers regarding appraisal decisions about research data has not yet been developed, though we have had initial discussions with the new Digital Archivist about it.
- Image-related services were not funded by the DRS programme and were not pursued.

E. Research Data Support

25) Refresh Communications Plan to ensure key messages are getting out to all stakeholders in line with DRS objectives.

Priority: High

- A new 'working' Communications Plan written in 2018 including key messages, stakeholder analysis, communication channels within team and beyond, as well as a new academic calendar highlighting when in the cyclical year tasks need to be done.

26) Vibrant network of data managers and curators speaking about their use of data with service representatives and each other.

- Two Dealing with Data annual events were held during the period, with a theme on open science and on collaborating with data. Each event had a call for papers and ~100 attendees.
- The Research Data Workshop Series was held in Spring 2019, on the topics of DataVault, Sensitive data, ELNs. These attracted data managers and support staff as well as academics.
- The team has engaged with the Edinburgh Open Research Champions group, which sprung up in the BioSciences.
- A Research Outputs Forum was set up in collaboration with colleagues in Library Research Support to inform research administrators about research data service developments, REF and Open Access. A second meeting is due to be held this month.

Priority: High



27) Forward-looking training on data protection in light of GDPR.

- The team contributed to the Data Protection for Researchers course developed by the DPO.
- “Data Protection Do’s and Don’ts” prepared as a Quick Guide.
- Revamp of Working with Personal and Sensitive Data course to align with university policy and law. This was trialled at a Portuguese university by invitation, for consultancy income.
- Revamp of MANTRA course in light of GDPR, open science and FAIR principles.

Priority: High

28) Conduct privacy impact assessments on service components and/or service as a whole.

Priority: High

- PIAs completed on the new service components, Data Safe Haven and DataVault.
- Privacy statements issued for each service component online.

29) Training on and offline is responsive to user needs, new tools, open science / digital scholarship new norms for transparency and reproducibility.

Priority: Medium

- Work completed to modularise and update all training session materials to allow easy customisation for different groups and timeframes.
- MANTRA overhaul half-way complete to modernise the material, maintain reputation.
- Cleaning Data with OpenRefine added to hands-on training sessions (on hold during COVID working from home).

30) Ensure researchers have access to expert advice in anonymising and pseudonymising their data.

Priority: Medium

- The team is trained sufficiently to check submissions for personal data disclosure.
- The team provides advice on data disclosure in helpdesk enquiries.
- A training course has been piloted using a data disclosure tool (on hold during COVID working from home).
- Gatekeeping procedures are developed on a per project basis in Data Safe Haven. These are then audited by the team for compliance.

31) Alignment with College and School Ethics committees for data-related guidance.

Priority: Medium

- Engagement with ERO around research integrity including participating in training for PGRs.
- Participation in the REIRG group rationalising the ethics process across the University.
- Cross-referrals between ourselves and school ethics committees when assisting users.

32) Revisit statistical training and consultancy offerings with relevant parts of IS in light of Innovation Project outcome (Scoping Statistical Support).

Priority: Low

Known gaps for Research Data Support



- A permanent network of data managers has not been established by the service. A Community of Edinburgh Research Software Engineers (CERSE) was however set up by others.
- There is no data disclosure direct service provided to research groups but advice on general information governance is provided (and improving as our knowledge grows).
- Statistical training and consultancy is not offered, though there is now Q-Step and International Centre for Mathematical Sciences which do. SPSS training course remains in our portfolio, with an emphasis on using it to produce reproducible research.