Edinburgh Research Explorer

A Decade of Research Data Management at the University of Edinburgh:

OPEN ACCESS

A Decade of Research Data Management at the University of Edinburgh: Looking Back, Looking Forward

Robin Rice

 https://orcid.org/0000-0003-0214-6559

University of Edinburgh, UK

ABSTRACT

The chapter examines how research data management (RDM) policy and service developed at the University of Edinburgh, comprising three sections. First, the current shape of the service is examined, covering all of the service components and their utility to researchers in turn. Second, the service beginnings and foundations are laid out, such as the 2011 university RDM Policy and how this represented a strong institutional commitment to RDM, and how the first RDM Roadmap addressed policy fulfilment through service rollout. A look at governance, funding, and organisational structure completes this contextual background. Third, strategic directions are explored, focusing on the gaps and achievements of the most recent completed roadmap and a preview of a newly modernised RDM policy and its upcoming implementation.

Keywords: Research Data Support, University Policies, Information Services, Library and University Collections, Research Data Lifecycle, Digital Repositories, DataShare, Service Governance, RDM Training, MANTRA

INTRODUCTION

Research Data Management (RDM) refers to good practice in planning, collecting, storing, documenting, sharing and preserving the data generated and used in research projects. Good data management practice offers a number of benefits to researchers and the public, and an obligation of the research institution is to provide support and infrastructure for its researchers' data.

Both research funders and publishers increasingly expect that data resulting from research projects should be made available for scrutiny and reuse, where legal and ethical requirements allow. Using appropriate standards, technology and trusted services to manage and share their data helps researchers to make their data FAIR (findable, accessible, interoperable, reusable). In following the FAIR principles (GO FAIR, n.d.), researchers are improving the public return on investment in science and research by adopting open science or open research practices.

The University of Edinburgh, formed in 1583, is an internationally performing public research university with more than 40,000 students, nearly half from outside of the UK. Most academic subjects are covered, with 21 schools across three colleges: Arts, Humanities & Social Science; Medicine and Veterinary Medicine; and Science and Engineering. A central Information Services Group provides Library and IT services to the staff and students. The university is a member of several associations of research-intensive universities, including the Coimbra Group, League of European Research Universities, Russell Group, Una Europa, and Universitas 21. The university aspires to help organisations across 10 industrial sectors, for Edinburgh to become "the data capital of Europe," through the decade-long government-funded Data Driven Innovation programme (University of Edinburgh, 2022a).

The University of Edinburgh has a relatively long history of supporting researchers in research data management, through investment in innovative and collaborative service options, training to develop the skills and capacity of practitioners, and policy development. This chapter will delve into this rich history, examining successful innovations and ongoing challenges on the journey of the Library and University Collections as a partner with fellow service providers, and with the academics it serves, in building data science ecosystems.

Edinburgh's RDM development milestones will be visited in the section, "Service beginnings and foundations," after a look at the "Current shape of the service". The third and final section, "Strategic directions – recent past, present and future" will indicate upcoming intentions, including a renewal of the RDM policy.

CURRENT SHAPE OF THE SERVICE

In trying to avoid acronyms for particular service functions, Edinburgh has found itself coining ever more portmanteau names starting with 'Data…' in the tradition of DataShare and DataStore as it has built up its offerings. The wisdom of this route has been questioned internally, not least by seasoned service providers who inadvertently mix up the names themselves, but it has been tricky to find a simple way out of the situation. While trying to avoid throwing these names at newcomers to the service (they are not the titles of service web pages, but rather a plain English description is used for each), the team mostly laughingly embraces the proliferation. There is now DataStore, DataSync, DataShare, DataVault; the Data Safe Haven, which launched in 2019, was closed only two years later, though not because it broke the naming rule, as will be explained under Strategic Directions. Figure 1 is used to represent the service as a whole in promotional materials, indicating the centrality of the researcher's data regardless of which of the several tools may be utilised.

Figure 1. Researcher's data at the heart of the service

Source: University of Edinburgh (2019)
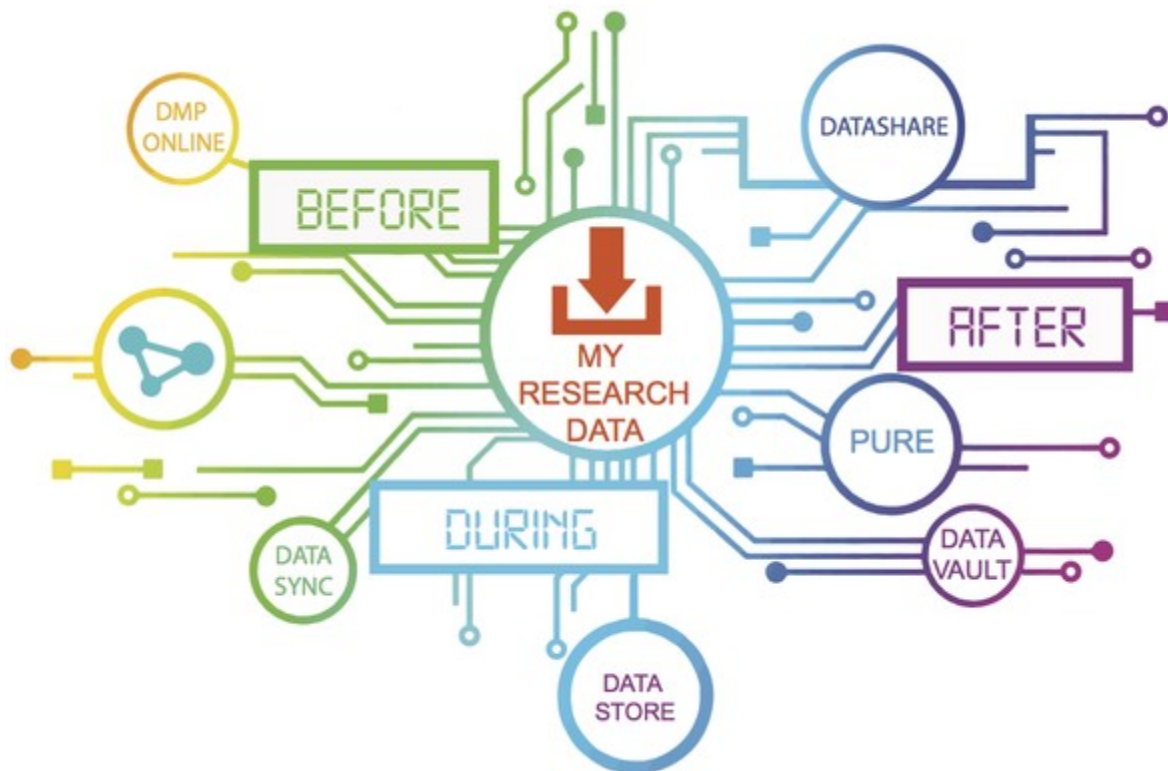
The website and other promotional materials tend to convey the service components in terms of a data or research lifecycle, overlaying the service functions onto the research tasks at each stage of the cycle. While lifecycles are useful for gap analyses or learning tools in an RDM context, they can be a barrier for researchers quickly trying to find the right tool for the job (Rice and Southall, 2016, pp. 76-80). For this reason, the service has adopted two designed images that aim to simplify the lifecycle so it can be understood at a glance. The first (Figure 2) is used on web pages and promotional materials as a symbol for data management planning. The second (Figure 3), allows grouping of service components by colour coding, and utilises hands of a clock to show stages of the lifecycle. The four categories of tools and support that make up of the service are: before; during; after; training and support.

Figure 2. Simplified data lifecycle

Source: University of Edinburgh (2019)

Figure 3. Introduction from the Research Data Service brochure

Source: University of Edinburgh (2019)

Research Data Service

We provide and advise upon a range of tools and support to help you before, during and after your research project:

BEFORE: identify existing datasets for reuse, plan how you will collect, manage, and store your data, including storage and backup, sensitivity of the data, and whether it can be shared;

DURING: analyse data, store and backup active data, synchronise your data across devices and with your research collaborators, and version your software and files;

AFTER: record, share, and archive your data for the long-term;

TRAINING AND SUPPORT: online and face-to-face workshops and courses, plus direct support when you need it.

BEFORE	DURING	AFTER	TRAINING & SUPPORT
create a data management plan	working with data	share and archive your data	
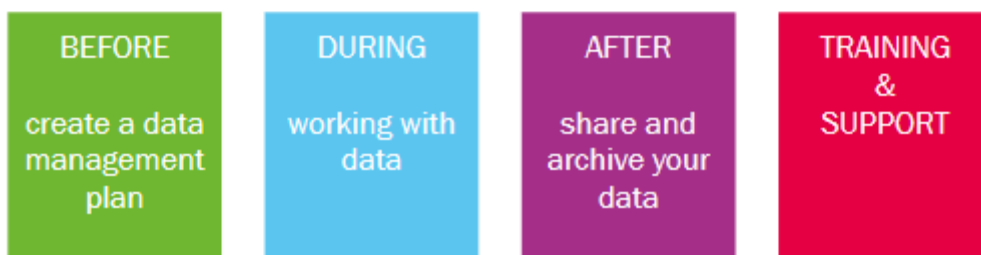
Accordingly, the rest of this section will cover the tools and support offered under each of the four categories. It is interesting to note that the logic of the simplified lifecycle – before, during, after the research project – has been utilised to frame some other European institutions' RDM programmes as well (Hettne, et al., 2020; Töwe & Barillari, 2020.)

Tools and Support: Before You Begin

"Create a data management plan to ensure that your research proposal is as strong as it can be," (University of Edinburgh, 2021a).

- 1. **DMPOnline**: a tool provided by the Digital Curation Centre and customised by the university's research data support team through a subscription model. It offers templates and guidance for numerous UK and international funders' requirements updated by the DCC. The university template is for those without a funder's template to follow – students and unfunded projects - with guidance specific to university service provision.
- 2. **Funder data management requirements**: a web page summarising the university's primary funding organisations' data management and sharing requirements with links to their policies. This is updated on an ongoing basis..
- 3. **Sample data management plans**: kept behind a university login, the service is acquiring current plans that have been written as part of successful bids in different disciplines.
- 4. **Rapid response advisory service:** for those unsure of how to write a plan, the team offers direct support and advice. This can sometimes be quite in-depth with verbal consultations and iterative versions, depending on the complexity and amount of time the researcher has before the due date. In some cases, requests arise on the back of a training session; in other cases, school-based research support staff may wish to escalate a tricky RDM problem to an expert.
- 5. **DPIA advice and support**: Sometimes service staff need to advise on a data protection impact assessment (DPIA), which is required by the university's Data Protection Officer and Ethics Committees for projects involving collection, receipt or transfer of personal data.
- 6. **Scheduled and bespoke data management plan training:** *Writing a Data Management Plan for your Research* is a regularly scheduled interactive training session offered by the team, who can also tailor a bespoke session for a particular group, according to preferred content, discipline, and time length. A 'Train the trainers' approach has been taken to spread expertise and raise capacity for supporting writing plans to staff working in the research office and those providing research support in schools (the University of Edinburgh equivalent of faculty departments).

Tools and Support: Research in Progress [During]

"Tools to help store, manage, protect and collaborate on research data during your project," (University of Edinburgh, 2021b).

- 1. **Find and reuse data**: support for discovering and accessing data for secondary use in research, learning and teaching. Before the Research Data Service, data librarians managed a catalogue of datasets available from national and international data producers. In 2020 it was decided this was too cumbersome to keep current, and searching for datasets directly over the internet was becoming easier. It was paired down to a small set of links under 4 categories: dataset search engines (including Google Datasets); data providers UK; data providers international; and university resources (e.g. database subscriptions).

- 2. **Active data storage – DataStore**: a centrally managed file store offered to all research staff and postgraduate research students to store their data securely. It is backed up to 3 different locations with offsite disaster recovery platform and offsite backup and retention on tape. A 'free at point of use' individual allocation of 500 GB is standard for every academic staff member and postgraduate research student. Individual allocations can be pooled into group spaces for free; additional capacity for group spaces can be purchased (and is expected to be itemised on research grants). There is provision for very large allocations (petabyte-scale) through shared procurement. The management of group spaces is devolved to school-based computing officers through a web application called Storage Manager.
- 3. **Advice on working with sensitive data:** while DataStore is secure for most purposes, it is not encrypted and does not meet certain cybersecurity standards, so does not meet all data controllers' requirements for secure storage and handling of sensitive data. Service staff offer pointers to information security pages on the university website or direct advice on applying additional safeguards such as encryption, or finding a data safe haven to work in. Advice is offered in handling and transferring data safely, using encryption, permissions and other safeguards. A regularly delivered training session, *Working with Personal and Sensitive [Research] Data,* covers data protection legislation, ethics in terms of data handling, and using subject consent forms and information sheets, as well as the basics of anonymising qualitative and quantitative data (see Training sessions, below).
- 4. **Version control services:** allow users to store active documents, source code and other research objects. When these documents are updated with newer versions, the old versions are still kept, enabling the user to revert to an older version if needed. These services also allow for multiple people to collaborate on projects, by letting users check out a document to work on locally and allow management of multiple changes by independent users. The Research Data Service offers two options. Both repositories may be used by Edinburgh University researchers and their external collaborators, and are available at no charge, under a storage threshold of 10 GB.
    - **GitLab:** suitable for distributed development or where branching and merging is a common operation.
    - **Subversion:** is suitable for larger projects that require centralisation or granular access control. The IT Infrastructure team that runs these is looking to retire the older Subversion facility, once key projects are completed or able to migrate.
- 5. **Collaboration tools – DataSync:** a file-hosting service to synchronise and share research data with colleagues anywhere. It has an app to synchronise data to computers and mobile devices, and a web interface to allow access to data from any web browser. Data can be shared with anyone who has an email address, via the web interface (no account creation is required for those accessing stored data). Based on the open source software ownCloud, it is hosted on university infrastructure, and files up to 20 GB can be stored in the app, with the ability to point to a space on DataStore for larger file sharing requirements.

The service also recommends either the university wiki service (Confluence) or the university's licensed Microsoft Sharepoint for research collaboration, though they are operated by other teams.

- 6. **Open research tools:** These are cloud or locally hosted commercial tools in which the service has brokered licenses for university users or purchased an enterprise

subscription (RSpace and protocols.io). RSpace is an electronic lab notebook (ELN) originally invented by University of Edinburgh researchers; it is used by individuals and some research groups and labs from various disciplines in different ways. The protocols tool is predominately a secure (cloud) platform for developing and sharing reproducible methods. In both cases, in-depth tool support is carried out by the contracted companies rather than service staff. Computational notebooks are another type of tool and the service recommends Jupyter notebooks, as provided through a central service called Noteable: an open-source web application that allows users to create and share documents that contain live code from different languages, equations, visualisations and narrative text. Jupyter notebooks can provide a platform for open, reproducible science, or be used as a tool in learning and teaching.

Tools and Support: Approaching Completion [After]

"Archive, preserve and share your data towards the end of a project, and link datasets to your publications with a DOI," (University of Edinburgh, 2021c).

- 1. **DataShare Open Access Data Repository:** an online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh. Benefits of depositing include: preserving data for future use (even after leaving the institution); creation of a permanent record, persistent identifier (DOI or digital object identifier), and a suggested citation to make it easy for others to cite the work; data will be more discoverable through Google and other search engines, maximising visibility and impact; ongoing management of online data is assured by the service; compliance with funder and publisher (and university) requirements to make data underlying publications available.

DataShare is a flagship component of the Research Data Service, which has acquired over 3,000 unique datasets from university researchers since 2010 (in increasing numbers overall – the first year only nine were deposited). Using DSpace software, the repository is supported by the Digital Library team in Library and University Collections, with approximately one FTE software developer time spread over two people, and about one FTE on the core team dedicated to quality assuring deposits and providing user support. (Each FTE is contributed by more than one staff member who has other roles; this ensures constant coverage.)

- 2. **DataVault Long-term Retention Solution**: the purpose of DataVault is to provide data owners with a long-term, low-cost, immutable and secure storage solution for research data that are no longer active, or not intended for publication. The service enables principal investigators to a) comply with funder and university requirements to preserve research data for the long-term and b) confidently store data from a given research project for potential retrieval at a future date. Where DataShare is free at point of use for users, DataVault has charges for projects with data totalling over 100 GB. The charge is based on size plus number of years of retention. UK research funders often have a minimum period of retention required for data they have paid to produce. A public record and a DOI is assigned for each project's set of data deposited in the DataVault; this allows other researchers or members of the public to discover its existence and request access from the data owner.

DataVault stores three encrypted copies in different locations (two on tape in university data centres, one in the cloud) with integrity checking built-in, so that the data owner can retrieve their data with confidence until the end of the retention period (typically ten years). DataVault's workflow takes into account the possibility/likelihood that the original data owner will have left the university when the period of retention comes to an end. Each vault will be reviewed by representatives of the university in schools, colleges or the Library, acting as the data owner, to make decisions on disposal or further retention and curation. If kept, the vault contents become university data assets, though at the cost of further storage and curation.

- 3. **Data Asset Register**: University of Edinburgh Research Explorer is the public view of the university CRIS which links people, projects, research outputs, and datasets. All data deposited into either DataShare or DataVault are automatically added to the CRIS, and the Research Data Service recommends that researchers record datasets deposited elsewhere by creating a dataset record and linking it to their publications, though the record may not be approved by staff if the location is not deemed a suitable archive (for example, if the dataset's location is simply a server without any persistent identifier). Ideally, all archived research data that is part of the scholarly record would be recorded in the register.

Training and Support: Throughout the Data Lifecycle

Research Data Support staff are available to answer enquiries, or to meet by appointment (in person or through Microsoft Teams) to help resolve any research data management issues at the point of need. The staff also carry out a wide range of training and outreach activities to raise awareness of RDM and generate new demand for the service.

- 1. **Research Data Service website**: a one stop shop for information about research data management and related services available at the University of Edinburgh. In addition to service information described below, the web pages host links to the RDM Policy and Roadmap, topical Quick Guides, and videos, including a commissioned, animated video describing the service in under three minutes for new users.
- 2. **Helpdesk Support:** If the answer cannot be found on the website, or if in-depth help is needed, users may contact the Research Data Support team by email or logging an enquiry into the university's call management system. Any call reaching the Information Services Helpline may be passed to the team if front-line staff determine they are best placed to answer it. This neatly solves the problem of a user needing to know about a team or service before they can request assistance.
- 3. **Blog, Twitter, newsletters**: Edinburgh Datablog is the main instrument for the team to disseminate new service developments (https://libraryblogs.is.ed.ac.uk/datablog), and has been since May, 2013. Regular service users and stakeholders are encouraged to subscribe to the blog. Copy which is written for the blog can easily be repackaged for other purposes, such as Information Services newsletters. The team has managed a twitter account since February, 2019 (@ResearchdataUoE) and some individuals use Twitter on a regular basis for current awareness in the RDM field. A new development is an Open Research Newsletter, comprised of news from the three service teams in Library Research Support and from champions around the campus, such as the Edinburgh Open Research Initiative and the ReproducibiliTEA reading group.

- 4. **Events**: Depending on resources, the team strives to host events to help early career researchers network, and to put the spotlight on timely or tricky issues in RDM. Before the Coronavirus pandemic, the team held an annual event called *Dealing with Data*, which called for contributions from university researchers (presentations or posters) with compelling keynote speakers and lightning talks from Digital Research Services staff (Papadopoulou, 2019). Going forward, this is likely to become a hybrid event with a general focus on open research. Library Research Support have also hosted a number of external events in collaboration with Digital Curation Centre staff, including a string of *Repository Fringe* events for UK-based repository managers and developers, and the FORCE2019 international conference.

Smaller half-day workshops have been used by the team to bring together researchers and support staff across the university to discuss specific topics of concern, or to gain feedback on requirements for new services.

- 5. **Training sessions**: These are delivered on a regular basis (in person or by using Microsoft Teams) several times a year through Information Services' Digital Skills service. The following topics have been developed and delivered by the team over a number of years, and registration for these courses is available to staff and postgraduate research students for free through an online events system called MyEd.
    - a. *Realising the Benefits of Good Research Data Management:* interactive course, split into two 90 minute sessions. "This in-depth course will provide researchers at all stages of their career with an accessible and practical guide to Research Data Management (RDM) and how it can benefit their research. It covers the entire research data life cycle from data management planning through to improving the visibility and impact of research data in order to increase a researcher's profile. With practical hints and tips throughout, this course is ideal for those just setting out on a career in research as well as anyone looking to update or refresh their skills," (University of Edinburgh, 2021d). Learning objectives:
        - i. Recognize the importance of RDM and Data Management Planning.
        - ii. Understand the difference between sensitive and non-sensitive data and how that will impact on their research and RDM.
        - iii. Apply basic RDM skills to their daily research practices.
        - iv. Determine the extent to which their research data is Findable, Accessible, Interoperable, and Reusable (FAIR).
    - b. *Writing a Data Management Plan for your Research:* interactive course, two hours. "At the end of this practical workshop you will have produced a first draft Data Management Plan (DMP) for your research project. You will understand the basic components of a good DMP," (University of Edinburgh, 2021d). Learning objectives:
        - i. Understand the necessity/benefits of producing a DMP.
        - ii. Know how to register for and use DMPOnline.
        - iii. Draft a basic DMP to complete after the course.
    - c. *Working with Personal and Sensitive Data*: interactive course, two hours. Prerequisite: complete the university's mandatory, online Data Protection Training course. "Researchers today are pressured to share their research data and make it accessible to other researchers. But what if you have collected sensitive or confidential data? For many researchers, the sensitivity of research data is one of the main barriers to data sharing. Fear of violating ethical or

legal obligations, lack of knowledge about disclosure control and the time required to anonymise data to a suitable standard often prevent valuable datasets from seeing the light of day. In this two-hour awareness raising course, we introduce how to collect, share, store, and protect the sensitive data you may encounter as part of your work," (University of Edinburgh, 2021d). Learning objectives:

- i. understand legal and ethical issues involved in data collection, data management, and data sharing with regard to human subject research;
- ii. have knowledge of GDPR [the European Union's General Data Protection Regulation] and UK data protection legislation, and what these laws mean for research and research data;
- iii. learn strategies for working with personal and sensitive data that are both data protection-compliant and follow good practice in research integrity and research data management.

o d. *Archiving your Research Data* (new in 2021-22): interactive session, one and a half hours. "Learn how to approach the key decisions around archiving your research data, such as: which research data to archive and when; whether to release your data under an open licence; how to select the best repository for your data; how long to store your data after the end of your project; and how to determine costs of long-term storage," (University of Edinburgh, 2021d).

o e. *Various 'hands-on' training sessions:* conditional on access to training rooms, social distancing requirements, and available staff resource, the following courses are sometimes offered:

- i. Data Handling Using SPSS
- ii. Data Cleaning with OpenRefine
- iii. Introduction to Visualising Data in QGIS/ArcGIS
- iv. Assessing Data Quality in Quantitative Data
- v. Assessing Disclosure Risk in Quantitative Data
- vi. Edinburgh DataVault – supporting users archiving their research data

o f. Other university training channels - given available staff resource, the team sometimes delivers courses for the following training programmes: Edinburgh Carpentries; Centre for Data, Culture and Society; Institute for Academic Development; The Wellcome Trust Clinical Research Facility; Doctoral Training Centres; and the Library's Dissertation Festival. The team will generally fulfil requests to provide bespoke training for particular groups, by tailoring existing resources accordingly (to discipline, audience, time length).

- 6. **Online training**
  - o a. **Research Data MANTRA** was originally developed using open source software, Xerte Online Toolkits, with funding from the Jisc Managing Research Data Programme in 2010-11, which made it one of the first open, online learning resources in RDM. The resource has been lovingly updated by the team for both local and international audiences ever since (Rice, 2014). The site attracted over 27,000 sessions and over 15,000 unique users in academic year 2020-21. The highest proportion of those, geographically, was approximately 20% from the UK and 5% from the city of Edinburgh. Users have rated MANTRA 3.79 out of four.

"MANTRA is a free, online non-assessed course with guidelines to help you understand and reflect on how to manage the digital data you collect throughout your research. It has been crafted for the use of post-graduate students, early career researchers, and also information professionals. It is freely available on the web for anyone to explore on their own," (University of Edinburgh, 2021e). Eight self-paced modules cover the following topics, with a further ninth option to practice data handling skills in a particular software environment:

- i. Research Data in Context
- ii. Data Management Planning
- iii. Organising Data
- iv. File Formats and Transformation (soon to be republished as File Formats and Digital Preservation)
- v. Documentation, Metadata and Citation
- vi. Keeping Research Data Safe
- vii. Protecting Sensitive Data
- viii. FAIR Sharing and Access
- ix. Data Handling Tutorials: SPSS, R, ArcGIS, NVivo
    - b. **Research Data Management and Sharing MOOC**– this Massive, Open, Online Course developed in 2016 by the University of North Carolina, led by Dr. Helen Tibbo, includes a pragmatic *Working with Data* section loosely based on MANTRA, presented on the team's behalf by a staff member of the Digital Curation Centre. Over 27,000 learners have enrolled in a 5-week cohort and about 3,000 have completed the 14 hour course so far. Users have rated the Coursera MOOC 4.7 out of five.

"The MOOC uses the Coursera on-demand format to provide short, video-based lessons and assessments across a five-week period, but learners can proceed at their own pace. Although no formal credit is assigned for the MOOC, Statements of Accomplishment will be available to any learner who completes a course for a small fee," (The University of North Carolina at Chapel Hill & The University of Edinburgh,, n.d.). Subjects covered in the 5-week course follow the stages of any research project. They are:

- i. Understanding Research Data
- ii. Data Management Planning
- iii. Working with Data
- iv. Sharing Data
- v. Archiving Data

Incidentally, members of the team also appear in some videos in the Research Data Management Services MOOC created by the Digital Curation Centre and Research Data Netherlands in 2019 explaining concepts to potential RDM service providers, such as how to develop an RDM Roadmap (Digital Curation Centre & Research Data Netherlands, n.d.).

- c. **Data Mindfulness, Making the Most of Your Dissertation:** the result of an Information Services funded 'innovation project' aimed at delivering basic RDM training to undergraduate honours students, who choose to complete an extra fourth year of their degree by conducting a piece of original research. The course, developed by a PhD student intern, focuses on aspects of writing a dissertation that first-time researchers seem to feel are not conveyed in the normal curriculum, such as record-keeping, file naming and organising information, as well as concepts of research

integrity, transparency and data citation. The original 2018-2019 course was delivered interactively in small groups, but the instructor left a legacy of a course handbook and a set of videos of herself delivering topical presentations on YouTube. This has since been converted to the university's Blackboard Learn platform, with written introductions and quizzes. A version of the course has been adopted by the Library for its *Libsmart II* Learn-based information literacy course, as well.

Service Usage and Spread

From the start, usage statistics have been meticulously measured and reported. Yet key performance indicators are difficult to identify, because of the problem of knowing what the ideal total take-up would be for such a relatively new and pioneering service. Published benchmarks from other institutions are not easy to come by, and when they are it can be hard to know if the services or populations served are similar enough to warrant comparison. Additionally, the university's RDM policy makes clear that use of external, domain-specific tools and support is acceptable or even advisable, so even a simplistic measure like counting use of a system against a denominator such as number of researchers or funded research projects does not give a full picture. Generally, internal stakeholders are interested in knowing how one college grouping compares to another, whether usage is increasing overall, or compares to the same quarter in previous years, or whether a particular school uses particular functions of the service.

Meanwhile, the following captured headline usage statistics from the calendar year 2020 are shared. (While 2020 was not a typical year because of the COVID-19 pandemic, none of the statistics captured dropped more than might be expected under circumstances of enforced remote working and pandemic-related disruption. From the standpoint of early 2022, 2019 remains a high-water mark overall, which also marks the period when the teams were most fully staffed.)

- **Data Management Plans:** 377 plans were created in the University's DMPOnline tool. (Other plans are likely to have been written without the tool; this is not currently captured.)
- **Terabytes Stored in DataStore:** As of December, 2020, Medicine and Veterinary Medicine was storing 4,787 (of which the Institute of Genetics and Cancer has 3,278); Science and Engineering 1,992; Arts, Humanities and Social Sciences 414.
- **RDM Support Calls**: The Research Data Support team logged 513 calls through the University's helpdesk system. About half were to do with data archiving (including support for deposits); the remainder were evenly spread for supporting DMPs; sensitive data enquiries; and non-categorised.
- **Open Data Deposits:** Two hundred twenty-one items were deposited by university researchers in the open access data repository. Since the repository began in 2008, over 3,000 items have been deposited, from all the 21 schools except one (Business).
- **RDM Training Sessions:** 1,196 people attended training or awareness sessions, 469 of whom were research students. (This includes sessions where RDM staff were invited to participate, including the popular Carpentries courses, which was conducted online in 2020.)

SERVICE BEGINNINGS AND FOUNDATIONS

A number of antecedents have helped to make RDM a strength at the University of Edinburgh. Edinburgh University Data Library, formed in 1984, bridged local computing and library support by providing expert support and access for machine-readable datasets for analysis. In 1996, EDINA (not an acronym but an homage to a Robert Burns poem and nickname for Edinburgh) was formed from an initial grant to act as a Jisc-funded data centre providing online services for UK Higher Education. The National e-Science Centre based at the Universities of Edinburgh and Glasgow (2002-2011) was a powerhouse for development and growth of national and European grid computing and data driven research. Actors from these organisations worked together with UK partners to create the Digital Curation Centre as a centre of expertise headquartered in Edinburgh, focusing on both research and capacity building in digital curation, in 2005. Meanwhile, the Digital Library section of the Edinburgh University Library developed local expertise in digital preservation and repositories, while the Unix Group of Edinburgh University Computing Service built local research computing systems, and the Edinburgh Parallel Computing Centre became a national supercomputing service centre.

RDM Policy

There is no difficulty putting a date on the beginning of the research data management (RDM) programme at the University of Edinburgh. Without a doubt, the first University Research Data Management Policy was put into effect in May, 2011. The policy was a first not only for Edinburgh University, it was seemingly the first RDM Policy of any UK university, though precedents were found in Australia, including Monash University, (Jones, 2013). In the wake of UK research funders rolling out mandates for data sharing and data management planning, the ten-point policy rapidly become a template for other UK universities to fashion their own approach to RDM policy.

This was hastened by the second Managing Research Data programme by the UK Higher Education funder Jisc, which focused on helping research institutions – often led from their libraries – to develop data policies and infrastructure as a response to increasing RDM requirements by UK research funders. (Jisc was founded in 1993 as the 'Joint Information Systems Committee,' but the acronym was dropped in 2012. The organisation provides digital solutions for UK education and research.) The Digital Curation Centre was funded by Jisc to provide general guidance and specific consultancy to several higher education institutions on RDM policy and infrastructure development as part of the Managing Research Data programme (Davidson, et al., 2014, pp. 216-217).

Early Institutional Commitment

At Edinburgh the "aspirational" policy was simultaneously matched with a capital investment to purchase hardware infrastructure to support the aims of the policy and provide every staff and student researcher with a half a terabyte of centrally managed, backed-up storage space. A Research Computing Survey of academic staff carried out in 1997 had already illuminated the dire need for reliable storage space by researchers (Ekmekcioglu, 2010), so this was a welcome development. Additionally, data librarians at the university had established a proof of concept institutional data repository using DSpace, as a deliverable for the externally funded DataShare project in 2008. When the Research Data Management Programme's steering group was formed in 2012, it already had the lynchpins of the service in place:

DataStore, for working with active data, and DataShare, for archiving and sharing the final versions of research data.

Further evidence about requirements for the nature of support needed by researchers at the university was revealed in the gaps identified by the Data Audit Framework, University of Edinburgh pilot case study, as summarised in Rice & Haywood (2011, pp. 4-5) and Rice & Southall (2016, p. 72):

- Storage provision is often insufficient.
- Data value is perceived as high and long retention periods needed.
- A lack of a formal data management plan; ad-hoc practices.
- A lack of guidelines and standardised procedures in creating and storing data.
- Minimal metadata; much effort is expended in finding extant data on servers.

These insights were then translated into recommendations by the Data Audit Framework Steering Committee, namely guidelines, training, policy and further gap analysis. The framework, later renamed Data Asset Framework (DAF), is a methodology developed by the Digital Curation Centre, modified based on the experiences of the pilot projects.

RDM Roadmap

The first RDM Roadmap, therefore, was written to both realise the full potential of the policy, and to fill in the gaps, such as those found in the data audit report. The Roadmap was a high-level development plan, which articulated goals and indicated milestones, actions and deliverables to achieve them over an 18 month period (2012-2013). The initial goals were described in a paper by the support staff involved (Rice et al., 2013) and are listed under their categories below:

- 1. **Data Management Planning:**
    - a. Tailored DMP assistance for PIs submitting research proposals
    - b. Customise DMPOnline for optimal University of Edinburgh use
- 2. **Active Data Infrastructure**
    - a. To provide a globally accessible cross platform file store with sufficient capacity to satisfy majority of researcher use cases
    - b. Provide additional data access mechanisms to better support mobile devices and external collaboration
    - c. To provide mechanisms to address backup and synchronisation of mobile devices.
    - d. Provide a service to ensure integrity and long-term retention of golden copy research data.
- 3. **Data Stewardship**
    - a. To develop the data repository for enhanced deposit and discovery of data collections generated by university researchers
    - b. To provide a registry of research data assets in support of the university RDM policy
    - c. To ensure efficient interoperation between all RDM systems as well as Pure [the university's CRIS, or campus research information system]
    - d. To provide continuity of access for data assets with long term value
- 4. **Data Management Support**

- o a. Raise awareness of university and funder policies, and advocate the use of data management plans for all research projects
- o b. Create and revise [Information Services] data management guidance
- o c. Maintain, develop and promote online training modules
- o d. Create tailored, on-demand training for research groups and professionals
- o e. Trial an in-depth data management consultancy service

The goals resulted in a number of successful outcomes, though in some cases took much longer to fully achieve. (For example, the long-term retention digital archive envisaged in 2,d, above, was developed with scattered internal and external funding over several years, tested and made more robust, and finally was launched into service in early 2019, as DataVault). The RDM Roadmap became a living document for subsequent plans varying in length from 18 months to three years, and was useful to determine which activities required additional investment from the university. The Roadmap additionally formed the basis of an internal bid for such funds to cover necessary infrastructure and staffing.

Subsequent RDM Roadmaps and development funding filled in service components according to a gap analysis, using a data lifecycle approach which could be described as: tools and support needed before a research project begins or in the early phases; requirements during the active phase of a research project; and that which is needed to steward the data after the end of the project. Training and outreach was the fourth category in the roadmap, which focused on improving researchers' skills and knowledge and contributing to academic culture change, in terms of understanding the importance and value of good practice in RDM.

Governance

The university was fortunate to have high level champions for research data management, including the CIO and Librarian to the University, Professor Jeff Haywood, and more recently Mr. Gavin McLachlan. It was the CIO's wish in 2016 that the loosely connected activities and systems under the RDM umbrella became formalised as the Research Data Service, at which point the partnership between the Library and the IT Infrastructure divisions of the university's Information Services was solidified. A service owner and service operations manager were established as roles for existing staff to be accountable for the service; these mirrored the front-end and back-end teams running the service from the Library and IT Infrastructure respectively. The academic chair of the governing steering group became the Business Service Owner in accordance with the university's service management framework (the BSO being a formal proxy for the user community to be involved in decision-making for the service).

In 2018 the steering group was changed to an oversight group reporting into a new Digital Research Services (DRS) Steering Group. The DRS programme as a whole involves not just Information Services, but strives to be inclusive of all university service providers whether their remit is local, national or international. Ultimately, this is meant to ensure that service users are matched to the best existing solution for their research requirements. In this way, EDINA, the Digital Curation Centre and the Edinburgh Parallel Computing Centre, each of which might have a solution that Information Services do not offer, are included in the portfolio of services offered to researchers. In practice, researchers may choose to use internal or external services to help them carry out their research.

Also in 2018 the current Business Service Owner, Professor Andrew Millar, a research group leader in the School of Biosciences, formalised the Terms of Reference of the group, and made changes to its membership. Primarily, he asked the more senior managers of Information Services who had habitually attended (such as the Vice-Principal and Chief Information Officer, and Librarian to the University; the Library Director; Deputy Library Director; and Chief Information Security Officer) to remove themselves from the group, in order that the IT College Liaisons and academic representatives of the three colleges could better set the direction of travel for the service. It is noteworthy that while many academics may see committee positions as nothing more than bureaucratic necessities, Professor Millar has included this role on his research profile, demonstrating his belief in RDM being at the heart of robust research.

Funding

The DRS steering group, along with its other business, prioritises and recommends development projects and activities from an annual envelope of revenue and capital funds for strategic investment to its parent, the Knowledge Strategy Committee. In addition to core funding from the various divisions of Information Services for ongoing staff roles established in 2016 or earlier, this source of funding has been the life blood of development activity for the service. It has allowed it to expand and improve where necessary, utilising short-term staffing resources from other Information Services divisions or contracting external staff, such as project managers or software developers, through procurement-cleared channels for hiring consultants for fixed terms.

It is a condition of DRS funding that service teams strive to recover their costs, once the service becomes sustainable. For some types of tools, free-at-point-of-use service is the norm (such as minimal active storage space, use of the open access data repository). In other cases, income can be recovered from most major UK funders for RDM costs, such as large-scale storage or licences for electronic lab notebook software. This requires that principle investigators (PIs) include accurate costs in their research proposals, and that through either overheads or charging mechanisms, the funding ends up where it can be spent on sustaining the service, which is not always the case.

Organisational Structure

RDM became a new service focus of the Library, which added the post of Research Data Management Coordinator to its Library Research Support section in 2014, joining a senior librarian who had worked as an information technologist before becoming involved with the set-up of the RDM programme. The Data Library staff moved from EDINA to Library and University Collections (L&UC) in 2018 to form the merged Research Data Support team. The new team comprised just over six full-time equivalents, alongside two related teams of fellow librarians working in scholarly communications and research information systems, together forming Library Research Support. While the support team largely focuses on front-end enquiries, as well as data archiving, the service systems were built and are managed by two other teams in Information Services. Staff of the Research Services section of the IT Infrastructure division, (which evolved from the Edinburgh University Computing Service) are experienced at providing research data storage and high performance computing platforms for researchers, and provide technical expertise and support for both the Research Data Service and the Research Computing Service. The Digital Library development team of

L&UC, which are experienced at provision of library technology including digital repositories, develop and maintain DataShare and DataVault, the two data archiving platforms of the service.

The organisational move to lead RDM services from the Library coincided with a shift for the data librarians, from focusing on helping researchers access and use research data collections, to helping meet new demand for solutions arising from unprecedented creation of research data. This in turn, reflected change more generally in the research environment, from data scarcity to data abundance. It also matched a broader trend for European academic libraries to recognise the importance of developing RDM services *as a core library service.* As noted in a Europe-wide survey, "Library directors, whether or not their library offers RDS [research data services], strongly agree that research data stewardship is important, losing data jeopardizes future scholarship, and the library needs to offer RDS to remain relevant," (Tenopir, et al., 2017).

Staff Skills

An advantage for the Research Data Support team being based in the Library, organisationally, is that much of the work needing to be done in RDM support is commensurate with what academic public service librarians already do, if somewhat more specialised. These include: marketing services; raising awareness and advocating for culture change; educating about issues of data literacy and publishing; answering queries situated in the context of the learner or researcher; and helping to solve research-related problems through knowledge of metadata and online tools. As professionals, RDM staff can both learn from and influence librarians working in other domains, in the course of everyday work and meetings. In practice, some of the staff have library-related backgrounds and others have research-related backgrounds. This mix of skills helps the team to engage confidently with senior researchers and early career researchers alike, offering some depth of understanding to research projects' outputs, while incorporating a broader view about scholarly communications and open science: for example, the importance of the role of research data and code in the evolving scholarly record.

A literature review of skills needed to provide Library-based research data services does not tend to reveal a consensus view; confusion may be a better reflection. This perhaps reflects the ad hoc way that these services have arisen in particular settings (Ribas Semeler, et al., 2017, p. 6). It can also be difficult to pinpoint skills needed when resource availability means that there can only be one dedicated post, given the potential range of skills involved. A particularly astute piece of research (Kennan, 2016, pp. 4-7), about skills required for research data management based on both a literature review and practitioner interviews found that skills sets fell into 6 major categories, which fairly reflects the University of Edinburgh experience:

- 1. Interpersonal skills and behavioural characteristics.
- 2. Contextual knowledge.
- 3. Data specific knowledge and skills.
- 4. Information technology.
- 5. Training and advocacy.
- 6. Teamwork.

Roadmap Review: 2017-2020

The third version of the RDM Roadmap may have been the most ambitious to date, including 32 objectives falling under five main categories, four of which map to the general shape of the service portfolio described above, with a mixture of service-specific goals and goals for the university community: unification of the service; data management planning; active data infrastructure; data stewardship; and data management support. A few objectives that required development funding and were not funded by the Digital Research Services Steering Group were dropped, since they were, in effect, deprioritised.

Much was accomplished in this 3-year RDM Roadmap. Major development effort resulted in the eventual launch of two service components that had been identified as filling a prior gap in the data lifecycle: DataVault for long-term retention of research data not otherwise published as open access; and Data Safe Haven, for providing active research environments that meet data owners' requirements for advanced security. A huge amount of engagement with university users and stakeholders took place, as well as with external peers, resulting in greater shared understanding of the utility of the services offered and the evolving needs of researchers. However, with insight from three consecutive roadmaps, some of the more difficult problems around interoperability and research culture change proved to be sticky, or unable to be fully resolved from one roadmap to another. A closer, selective look follows.

Unification of the service: "Actions needed to achieve a single, unified and comprehensive Research Data Service, a key part of the Digital Research Services (DRS) programme," (University of Edinburgh, 2017). This category was in part to clean up the flotsam and jetsam created by the distributed Research Data Management Programme that preceded the service designation, and in part to identify efficiency savings by simplifying the service components or introducing better interoperability between systems.

Three of the five objectives here addressed cosmetic or communicative requirements – creating a common look and feel to the service components, offering a simplified user journey starting from the service home page and creating a coherency to the service as a whole. These were largely achieved through a revamp and user testing exercise on the website; deep linking from the Digital Research Services website; a printed and PDF service brochure completed by a graphic designer; a professional promotional video coupled with video case study interviews with academic users; and a new series of Quick Guides giving a user-centred, as opposed to supplier-centred, view onto RDM.

Interoperability was more challenging to achieve but some headway was made. A change to 'Tier 2' storage was made to allow better capacity for data to be close to the compute facilities, e.g. the Research Computing Service. This helped certain workflows, particularly for biomedical researchers. DataShare interoperability was improved with automatic ingest of records to the CRIS plus harvesting was set up or resolved in Google Dataset Search, Google Scholar, DataCite, Altmetrics, and Web of Knowledge Data Citation Index. An end-to-end service was not achieved, with continuation of a number of systems to address different aspects of the data lifecycle. This predilection for home-grown service provision in University of Edinburgh's RDM 'bundle' was described in an OCLC report as follows:

The terms "locally sourced" and "externally sourced" are general labels that obscure a great deal of nuance. An internally sourced RDM solution is not necessarily one that was built entirely from scratch—it could instead have been assembled from various components acquired elsewhere, such as open source applications, that are then integrated and adapted or customized to meet local requirements. Indeed, one of our case study partners— Edinburgh—adopted this approach in the development of some of its internally sourced RDM services (Bryant, et al., 2018).

Data Management Planning: "Support and services for planning activities typically performed before research data is collected / created," (University of Edinburgh, 2017). Two interrelated goals in this area have proved difficult to progress, perhaps unsurprisingly, as they involve dependencies on changes to research culture and behaviours.

- Increase the number of fully-costed plans written by researchers across schools, including university-funded research.
- Increase the number of schools/departments that require a simple DMP as part of a PhD research proposal (from zero).

The first goal was about increasing the total number of plans created as researchers adopted the practice as routine, but it was also about ensuring that full costs were included in research grants. The team could point to a few indicators of an overall increase in plans (more plans created in the DMPOnline tool; more requests to review plans logged.) But the team does not have oversight of all DMPs submitted with grant proposals, either to count them or to monitor for inclusion of costings; for example, they can be created in a word processor without using the tool. The team still needs to establish better processes for the timely sharing of information between Research Data Support and the Edinburgh Research Office about DMPs, as well as for extraction of information about the nature of costings in funded grants from research information systems effectively.

The second goal is about establishing an expectation or requirement on PhD research proposals. The target was met, in that a handful of schools have requested DMP training for PhD students; and an expectation, if not a hard requirement, exists in at least two schools (GeoSciences and Business). However, given there are 21 schools in the university there is still much work to be done. A revision of the ten-year old policy is one lever identified to employ to help change the status quo.

Active data infrastructure: "Facilities to store data actively used in current research activities and provide access to that storage, and tools to assist in working with the data," (University of Edinburgh, 2017). There were fewer gaps on stated goals in this area, and one extraordinary achievement was made. Infrastructure refreshment has taken place on DataStore, for which cost recovery models have worked well. Storage prices have been benchmarked favourably, though there has been feedback that prices are still too high. In some cases institutes have organised hardware purchases through the central service, providing economies of scale. DataSync (OwnCloud) was upgraded. Gitlab was introduced, and is eventually expected to replace Subversion for code versioning and other aspects of code management. A local instance of Gitlab fits well with Data and Software Carpentry training offered by colleagues within the University to modernise ways of writing and sharing code.

A number of incremental steps toward uptake of electronic notebooks were taken, with RSpace being both promoted and monitored during its 2-year trial. At the end of the period, a business case was made to the business service owner and steering group to continue the arrangement and implement cost recovery on the individual licenses, to be managed by the Software Services team. Data portability and 'no vendor lock-in' were primary considerations for the tool, criteria which could also be applied to other candidate tools. Later, protocols.io was requested by a Bioscience researcher to be added to the portfolio; it demonstrated its value during the 2-year trial and has been extended indefinitely. Although it is useful for making one's protocols (or other workflows) public before conducting research, the enterprise version allows private protocols to be shared within a research group, which has also been popular. EDINA's Noteable platform for Jupyter Notebooks is also showcased on the service website as a computational notebook option, though the service was originally designed as a learning and teaching, rather than a research platform per se.

By far and away the most momentous achievement in this category was the development and launch of Information Services' Data Safe Haven to fulfil the deceptively simple sounding goal of "Provide secure setting for sensitive data and set up controls that meet ISO 27001 compliance and user needs." An enormous effort from a very small sub-team from both IT Infrastructure's Research Services and the Library's Research Data Support team, with significant investment in hardware infrastructure and expert consultancy, brought the trusted secure environment for research data to a soft launch at the annual Dealing with Data event in November 2018.

The Data Safe Haven (DSH) offered a secure storage space and a secure analytic environment appropriate for research projects working with different kinds of sensitive data. A layered architecture with its own firewall, isolated from other University network traffic and the internet, the DSH provided secure access for authorized users via two-factor authentication using YubiKeys and a VMware Horizon Client. Full ISO/IEC 27001:2013 standard certification was achieved for the DSH as a recognised ISMS (Information Security Management System) by UKAS (United Kingdom Accreditation Service) in December 2019, and was maintained in the 2020 audit cycle through careful adherence to a set of over 100 technical and operational protocols. The facility was approved by a number of external data providers, including National Health Service (NHS) bodies. Flexibility was seen as a primary advantage, with individual builds, or 'bubbles', for each research project, and the ability for 'gatekeeping' procedures to be defined on a project-by-project basis, depending on their requirements. The support team assisted research projects in setting up gatekeeping roles and rules, and audited activities to ensure compliance within the overall scope of the defined ISMS.

Achieving ongoing sustainability from charges against research grants, however, has not proven possible, given the expense of hardware, consultancy fees for achieving and maintaining certification, and levels of staff expertise required to run this type of facility. After the Roadmap period ended, and in the middle of a pandemic-induced budget crunch, staff attrition meant that university senior staff decided not to continue to invest in this facility, given other options opening up to local researchers through the Edinburgh International Data Facility (EIDF). The EIDF already provides a Scottish National Safe Haven through a partnership of the Edinburgh Parallel Computing Centre and Public Health Scotland. The EIDF has plans to build secure 'private zones' to achieve the type of gatekeeping flexibility and to enable advanced medical and other data science projects to be

carried out, for example through the use of Python code, which is very often disallowed from conventional safe settings environments.

The new challenge for Research Data Service staff is to work seamlessly with this service provider to refer users or customers to appropriate data safe havens and other secure services such as the DataLoch (which gets its name from the Scottish word for lake), while continuing to support unfunded research projects in applying more affordable safeguards, such as simple encryption methods. A lesson learned is that for trusted research environments to be sustainable, scale matters, and a single university's clientele for such a 'boutique' facility may not be sufficient in terms of cost recovery.

Data stewardship: "Tools and services to aid in the description, deposit, and continuity of access to completed research data outputs," (University of Edinburgh, 2017). There were seven objectives in this area, and five of them were found to be met in the review process. Most of these were to do with the launch and improvement of DataVault and the improvement of DataShare. Additionally, Research Services has begun exploring service delivery for an imaging and metadata tool for microscopy (Omero), through a partnership with a school at the university who have deep requirements for the use of it. Four key objectives in this area will be examined below:

- Alternative to open sharing exists with immutable, replicated secure long-term storage.
- University repository that provides FAIR (findable, accessible, interoperable, reusable) principles in line with emerging funder and international principles
- Repository makes best use of metadata standards and citations to improve impact for researchers.
- Repository continues to meet trusted digital repository requirements for researchers and actively migrates supported file formats for long-term preservation.

Regarding the first, the DataVault facility was successfully launched in January 2019, following stages of internally funded project development to take an externally funded prototype built in 2016 to suitable levels of functionality, including encryption, metadata capture, and set-up of role-driven permissions. Additional software developers and a project manager were deployed, using Jira issue and project tracking software and an Agile method. User documentation has been written along with how-to videos. Training for support staff in schools was rolled out, so they would be able to refer academic staff to the service as needed.

A review process for decision-making about whether to delete or extend retention periods for expired vaults was agreed. A web-based dashboard was implemented to ensure university management of data when the original data owner leaves the institution. (A chain of command model is used, in which the absence of the original PI or data owner from the institution triggers the next role in the chain to become the data owner, responsible for decisions about access and deletion.)

In Spring 2020, data migration from the 'interim DataVault', was completed, which provided additional testing of workflows and identification and fixing of bugs. In August 2020, performance and speed were enhanced to allow up to 10 TB sized deposits. (Any number of deposits can make up a vault, which is associated with a unique research project and grant ID, where relevant.) The DataVault was able to meet its income target in years 2019-20 and 2020-21.

The other objectives involved activity around DataShare, including development releases. During the period the repository was also migrated from EDINA to Digital Library infrastructure and upgraded to DSpace 6 (a major release). The team answered surveys, attended webinars, and conducted framework exercises to determine how 'FAIR' data in the repository were. Some limitations were part and parcel of being a multidisciplinary repository; for example, standard metadata is only applied at the study, not file level of items, as staff do not have the expertise to do more than basic quality checks on such a heterogeneous range of content. Others were overcome through the upgrade process; DSpace 6 offers a more robust REST-API, for example. Another release resulted in Altmetrics badges being added to DataShare items with one or more references. DataShare's metadata schema was updated to meet DataCite 3.2 metadata schema's minimal requirements. A case study on data reuse was written for OpenAire, which examined DataShare citations from Data Citation Index and Google Scholar (Rice, 2020). An application to Core Trust Seal for trusted digital repository status was made by the team during the roadmap period. (After two reviews and revisions the repository was certified in January, 2022 for a three-year period.) The team migrated known obsolete file formats in DataShare using a manual workflow in June, 2020. Talks are underway with Digital Library staff about automating this process using Archivematica in the future.

In terms of gaps, some desired features were not implemented in DataShare during the roadmap period due to the need to migrate servers and upgrade software. Others were postponed, as DSpace 7 was not yet released and was anticipated to have features that would make the interface more flexible and attractive. Features that were envisaged but did not get implemented in the timeframe of the roadmap included integrating the image standard IIIF (International Image Interoperability Framework), and enabling a preview of textual and tabular content before downloading. Also, ORCID implementation was trialled but postponed, until the metadata schema could be updated to incorporate a field for the ID to be stored, as part of a future new metadata release.

Data management support: "Awareness raising and advocacy, data management guidance and training," (University of Edinburgh, 2017). A new Communications Plan was created 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.

One goal was a "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 (the UK Research Excellence Framework) and Open Access.

Two training related goals were met, focusing on data protection in light of GDPR and the new UK Data Protection Act passing in 2018 and a goal that "training on and offline is responsive to user needs, new tools, open science / digital scholarship new norms for transparency and reproducibility." Regarding the first, the team contributed to the Data

Protection for Researchers course developed by the University's Data Protection Officer. Also *Data Protection Do's and Don'ts* was prepared as a Quick Guide for the service website. Both the *Working with Personal and Sensitive Data* training session and the MANTRA unit, *Protecting Sensitive Data* were overhauled to align with the new legislation. Significant work was also completed to modularise and update all training session materials to allow easy customisation for different groups and timeframes. The rest of MANTRA is receiving a major update on a rolling basis.

Regarding gaps in the support area, there is no data disclosure direct service provided to research groups but advice on general information governance is provided (and improving as the team's knowledge about the topic and resources in the university grows). Statistical training and consultancy is not offered, by the team or any team in Information Services unfortunately, though the School of Social and Political Science and the International Centre for Mathematical Sciences help to fill this gap. SPSS remains in the service training portfolio, with an emphasis on using it to produce reproducible research.

Overall 22 of the 32 objectives were deemed to have been met, with the others either partially met or not met. All but one of the high priorities were met. There is much to inform the next roadmap. One of the identified gaps was that the Research Data Management Policy became ten years old in 2021; it was agreed that a rewrite or refresh would be an opportunity to address a number of lessons learned from implementing the 2017-2020 roadmap.

Policy Refresh

A new RDM policy was drafted in early 2021, followed by a consultation period with key research committees and stakeholders. The new policy became active in January, 2022, along with a new Publications and Copyright Policy; implementation of both to be stewarded by Library Research Support.

The headings of the policy signify the important issues covered: RDM and Research Integrity; Responsibilities (including the researcher's and the institution's responsibilities); Research Data Management Plans; FAIR Data Sharing; Data Repositories; Rights in Research Data; and Review [of the policy], (University of Edinburgh, 2021f). A regular review period of five years will ensure that the policy stays at the forefront of research culture change and continues to challenge the University to make the most of its research data assets.

The inclusion of FAIR principles brings the policy in line with increasingly ambitious funder requirements on appropriate data sharing and matches the aspirations of the University's Open Research Roadmap, published on its website (University of Edinburgh, 2022b). The inclusion of software code in the policy allows a University focused on being a leader in AI and data science to treat both code and data as first-class research objects.

Through clearer explanation and linkage with related University policies such as the Data Protection Policy and the Information Security Policy, the new RDM policy is grounded in relevant context, which will make conveying the policy principles in training and outreach simpler. A definition of research data is given, and RDM is grounded in the context of the university commitment to research integrity.

It is increasingly important that the university recovers its research data management costs from funders in order to make services sustainable, and this policy seeks to ensure this happens through a greater focus on data management planning. Indeed all research projects will be expected to include a data management plan, not just those with funder requirements for DMPs.

The policy will help the university with its strategic goals through enabling researchers to make broader and more positive impacts in their scholarly communications through adherence to research integrity and data sharing best practice. The policy will underpin service development, training and outreach to make the Research Data Service fit for purpose in supporting researchers in their effective data management tasks. The new roadmap covers the same five main themes of the policy, to ensure activities are directed toward policy fulfilment. The review period ensures the policy will stay in step with positive improvements in research culture and making scholarship more transparent and impactful.

REFERENCES

Bryant, R., Lavoie, B., & Malpas, C. (2018). *Sourcing and scaling university RDM services. The Realities of Research Data Management, Part 4*. OCLC Research., doi:10.25333/C3QW7M

Davidson, J., Sarah Jones, S., Molloy, L., & Kejser, U.B. (2014). Emerging good practice in managing research data and research information within UK universities. *Procedia Computer Science,* 215-222. doi:10.1016/j.procs.2014.06.035

Digital Curation Centre & Research Data Netherlands. (n.d.). *Delivering research data management services.* Futurelearn. https://www.futurelearn.com/courses/delivering-research-data-management-services

Ekmekcioglu, C. (2010). *Research computing consultation, 2007* [Data set]. University of Edinburgh. Information Services. doi:10.7488/ds/89

GO FAIR. (n.d.). *FAIR principles.*https://www.go-fair.org/fair-principles

Hettne, K. M., Verhaar, P., Schultes, E., & Sesink, L. (2020). From FAIR leading practices to FAIR implementation and back: An inclusive approach to FAIR at Leiden University Libraries. *Data Science Journal*, *19*(40), 1–7. doi:10.5334/dsj-2020-040

Jones, S. (2013). Bringing it all together: A case study on the improvement of research data management at Monash University. *DCC RDM Services Case Studies*. https://www.dcc.ac.uk/resources/developing-rdm-services/improving-rdm-monash

Kennan, M. A. (2016). Data management: Knowledge and skills required in research, scientific and technical organisations. In *Proceedings of the World Library and Information Congress: IFLA General Conference and Assembly: IFLA WLIC 2016* (pp. 1-10). International Federation of Library Associations and Institutions.

Papadopoulou, E. (2019). 'Dealing with Data' Conference at the University of Edinburgh. In Engaging Researchers with Data Management: The Cookbook (pp. 70-73). Open Book Publishers. doi:10.11647/OBP.0185

Ribas Semeler, A., Pinto, A. L., & Frota Rozados, H. B. (2017). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*. Advance online publication. doi:10.1177/0961000617742465

Rice, R. (2014). Research Data MANTRA: A Labour of Love. *Journal of Escience Librarianship*, *3*(1). Advance online publication. doi:10.7191/jeslib.2014.1056

Rice, R. (2020). *Enabling data reuse: case study of an institutional repository, Edinburgh DataShare*. https://www.openaire.eu/blogs/enabling-data-reuse-case-study-of-an-institutional-repository-edinburgh-datashare

Rice, R., Ekmekcioglu, Ç., Haywood, J., Jones, S., Lewis, S., Macdonald, S., & Weir, T. (2013). Implementing the Research Data Management Policy: University of Edinburgh Roadmap. *International Journal of Digital Curation*, *8*(2), 194–204. Advance online publication. doi:10.2218/ijdc.v8i2.283

Rice, R., & Haywood, J. (2011). Research Data Management Initiatives at University of Edinburgh. *International Journal of Digital Curation*, *6*(2), 232–244. doi:10.2218/ijdc.v6i2.199

Rice, R., & Southall, J. (2016). *The Data librarian's handbook*. Facet. https://www.facetpublishing.co.uk/page/detail/the-data-librarian/?k=9781783300471

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R. J., & Allard, S. (2017). Research data services in European academic research libraries. *LIBER Quarterly*, *27*(1), 23–44. Advance online publication. doi:10.18352/lq.10180

Töwe, M., & Barillari, C. (2020). Who does what? – Research data management at ETH Zurich. *Data Science Journal*, *19*(36), 1–6. doi:10.5334/dsj-2020-036

University of Edinburgh. (2021a). *Before you begin.* https://www.ed.ac.uk/is/data-management-plan

University of Edinburgh. (2021b). *Research in progress.* https://edin.ac/2lGbKwe

University of Edinburgh. (2021c). *Approaching completion.* https://edin.ac/2UXHO0v

University of Edinburgh. (2021d). *Workshops and courses.* https://edin.ac/34JfRON

University of Edinburgh. (2021e). *MANTRA Research Data Management Training.* https://mantra.ed.ac.uk

University of Edinburgh. (2021f). *Research Data Management Policy.* https://www.ed.ac.uk/is/research-data-policy

University of Edinburgh. (2022a). *Data Driven Innovation*. https://ddi.ac.uk

University of Edinburgh. (2022b). *Open research.* https://edin.ac/3C6z8pn

University of North Carolina at Chapel Hill & University of Edinburgh. (n.d.). *Research Data Management and Sharing.* Coursera. https://www.coursera.org/learn/data-management

KEY TERMS AND DEFINITIONS

Data Management Plan (DMP): A 'living' document that outlines how data are to be handled both during a research project, and after the project is complete.

Data or Research Lifecycle: A set of actions taken on a data set or during a research project that corresponds to good practice and enables data reuse.

Electronic Laboratory Notebook (ELN): A software tool that in its most basic form replicates an interface much like a page in a paper lab notebook. Protocols, observations, notes, and other data may be entered and stored.

Gitlab: Similar to GitHub, available on the web, GitLab is a locally hosted open source repository manager which provides version control and lets teams collaborate on code.

Open Access [Data] Repository: A digital platform that holds research output [research data] and provides free, immediate and permanent access to research results for anyone to use, download and distribute.

Open Research: Sometimes associated with open research methods, however open research is a more inclusive term applying to all disciplines, equivalent to open science.

Open Science: Research conducted and published in an intentionally transparent manner. It may include one or more of the following: Open Access publication; open research data; open source software and code; open notebooks; open infrastructure; pre-registration of studies.

Research Data: Digital or analog information that is collected, observed, created, or reused to produce, validate and enrich research findings and conclusions.

Research Data Management (RDM): The organisation, storage and preservation of data created during a research project. It covers initial planning, day-to-day processes and long-term archiving and sharing, or deletion.

REST-API: An application programming interface that conforms to the design principles of REST, or representational state transfer architectural style. It allows programmatic access to content of a web system, rather than exclusive reliance on a web interface.

Roadmap: A strategic plan that defines goals or desired outcomes and includes the major steps or milestones needed to reach it.