IJDC | *Conference Paper*

A Country-level Case Study: On the Evolution of UK Institutional Research Data Services

Ruth Mallalieu
University of Oxford
ORCID: 0000-0002-0409-6199

Robin Rice
University of Edinburgh
ORCID: 0000-0003-0214-6559

Abstract

This paper examines milestones and unique service aspects of six research-intensive Higher Education Institutions' approaches to research data management policy and service, almost one decade on from their respective beginnings, based on findings from a 2024 internal benchmarking study conducted by the University of Oxford, which consisted of interviews with library-based research data management service providers at five peer UK institutions. Both similarities and differences are examined, and milestones are mapped against external events and policies in the RDM field. Future directions, and areas of convergence and divergence especially, will be explored across six institutions: the Universities of Cambridge, Edinburgh, Manchester and Oxford, Imperial College London, and University College London (UCL).

Submitted 30 June 2024 ~ *Accepted* 18 July 2025

Correspondence should be addressed to Ruth Mallalieu, Email: Ruth.Mallalieu@bodleian.ox.ac.uk

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

A six-month project commenced on 1 September 2023 to consolidate the Bodleian Libraries' Research Data Management (RDM) services, following on from a University-wide review at the University of Oxford, carried out in 2019–2020 (Chiarelli et al., 2022), and from further internal reviews to address service gaps and to enhance the services for researchers. An RDM planning group was set up and project strands were scoped. The findings made the case, and outlined a workplan, for a position of Head of Research Data Management, informed by a benchmarking exercise with five peer institutions. The exercise consisted of semi-structured online interviews, which took place through January 2024, using grouped themes and questions to elicit quantitative and qualitative data points. These institutions formed the COLIM Jisc Band 1 (at the time of the interviews)—most research-intensive—UK Higher Education Institutions (HEIs): Cambridge, Oxford, UCL, Imperial, and Manchester. Additionally, the exercise included Edinburgh, as it has a mature and expansive RDM offer.

The benchmarking exercise, therefore, formed a strand of the internal project and aimed to uncover and describe the facets and depth of RDM service provision in a typical research-intensive university library. The exercise gathered information across the following categories: governance; support team; long-term preservation and storage; training, advice and support; data services; tools, and Digital Humanities.

This paper has arisen as the result of a further collaboration between the Universities of Oxford and Edinburgh; in particular the IDCC conference theme of looking back and looking forward resonated with the authors. The authors have worked across the benchmarking categories, curating the richest and most topical themes in order to distil the current state of play into seven main trends:

- Governance, policy, staffing
- Long term preservation and storage
- Trusted Research Environments (TRE)
- RDM tools and data services
- Communication, training and events
- Data champions and data stewards
- Digital Humanities.

The interviews yielded a rich dataset (a summary of which is provided in Appendix 1), which can be mined to track development of RDM services at leading UK HEIs, mapped against fundamental milestones in the external RDM landscape, and the priorities and aspirations for future service development.

Drivers and Milestones in Early RDM Developments

Before interrogating the interview responses, some context about the early development of RDM services in the UK provides additional context to the national scene. By the time of the tenth anniversary of the International Digital Curation Conference in 2015, the most research-intensive UK Higher Education Institutions (HEIs) had formed at least some of the building blocks for research data management (RDM) services.

Drivers included: forward-looking UK funder policies which required Data Management Plans (DMPs) to be submitted with research proposals and gave an expectation of data sharing; strong open access (OA) repository and computing support in research-intensive institutions, and research offices focused on organising and showcasing research outputs because of the Research Excellence Framework (REF) — a national exercise taking place every seven years, which determines institutional funding allocations for university research.

National centres of expertise such as the UK Data Archive, the Natural Environment Research Council (NERC) data centres, the Digital Curation Centre, and the Digital Preservation Coalition also helped to spread knowledge of good practice and policies. Jisc, a non-profit organisation serving UK Further and Higher Education, poured funds and energy into institutional-based development projects and partnerships from about 2010–2016 through two 'Managing Research Data' programmes, and then a 'Research Data Spring' programme, which offered expert support from the Digital Curation Centre to institutions with funded projects.

Given other countries' trajectories into institutional RDM, it may be instructive 'looking back' at just some of the external drivers and key milestones in the very early stages of development leading up to the present day services.

Table 1. Timeline of selected key milestones and documents in early UK RDM landscape.

Year	Drivers/Milestones	Key Documents
2003	National Institutes of Health in USA issue data sharing guidance for its 2001 policy adding that applicants seeking $500,000 or more are "expected to include a data-sharing plan in their application stating how they will share the data or, if they cannot share the data, why not."[1] Hugely influential on UK and other funders.	
2004	Launch of Digital Curation Centre	From Data Deluge to Data Curation. (Lord, Macdonald, Lyon & Giaretta, 2004).
2005	First International Digital Curation Centre Conference: Bath, England. Dr Clifford Lynch, Director of the Coalition for Networked Information (CNI), provides the closing keynote address which becomes a time-honoured tradition.[2]	Digital Repositories Review (Heery and Anderson, 2005).
2006	The JISC Repositories and Preservation Programme funds HE institutional projects worth 14 million, over a three year period.[3]	

[1] https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html
[2] https://www.dcc.ac.uk/events/conferences/1st-international-digital-curation-conference/dcc-conference-programme/
[3] https://digital-scholarship.org/digitalkoans/2009/12/15/jisc-repositories-and-preservation-programme-final-evaluation-report/. [The actual report has not been located.]

Year	Drivers/Milestones	Key Documents
2007	Wellcome Trust publishes its Policy on Data Management and Sharing, two years after its Open Access policy for research publications (Carr, 2017). Research Data Management Forum (RDMF) is created by DCC, Research Information Network, Jisc, for UK data professionals to engage in workshop-style events. First event in Manchester.	OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007).
2008	The Arts and Humanities Research Council stopped funding their Arts and Humanities Data Service (except Archaeology Data Service). The third Open Repositories conference is hosted at Southampton University in April; the first Repository Fringe is hosted in Edinburgh in July.	RCUK drafts Common Principles on Research Data Policy, eventually superceded by 2016 Concordat document, (UKRI, 2016).
2009	'Climategate': East Anglia University's Climatic Research Unit has more than 1,000 e-mails hacked from its servers by climate skeptics. The University was later criticised for its handling of Freedom of Information (FOI) requests, spooking other UK university leaders into improving research transparency. DataCite, a multi-country initiative to make data citable by issuing special DOIs (digital object identifiers), is founded in London at the British Library.[4]	The Fourth Paradigm: Data-intensive Scientific Discovery, is published by Microsoft Research (Hey, Tansley & Tolle, 2009)
2010	Data Seal of Approval, developed by DANS in the Netherlands, pre-cursor to CoreTrustSeal, issues its first set of requirements for trustworthy digital repository certification.[5] Jisc funds 5 RDM training modules including University of Edinburgh's Research Data MANTRA, as part of Managing Research Data Programme.[6] The authors of the Panton Principles were honoured as SPARC Innovators, for their	

[4] https://datacite.org/blog/how-it-all-began/
[5] https://www.coretrustseal.org/about/history/data-seal-of-approval-synopsis-2008–2018/
[6] https://researchdata.jiscinvolve.org/wp/2010/08/01/new-projects-in-the-jisc-managing-research-data-programme/

Year	Drivers/Milestones	Key Documents
	2009 call to use open licenses to dedicate data underlying science to the public domain.[7]	
2011	Big Data is added to the Gartner Hype Cycle,[8] later peaking in 2013. Figshare repository is launched by PhD student at Imperial College London, supported by Digital Science from 2012.	UK Data Archive publishes Managing and Sharing Data: a Good Practice Guide for Researchers. (Eynden et al., 2011).
2012	DCC's DMPOnline tool launched as a free cloud service based on open source software. Later added institutional subscriptions and from 2018 cooperated with California Digital Library's DMPTool on a shared software roadmap.	Reference Model for an Open Archival Information System (OAIS) is updated. Nobody reads the document but everybody includes the graphic in their presentations (Consultative Committee for Space Data Systems, 2012). The Royal Society (2012) issues forward-looking Science as an Open Enterprise report advocating data sharing: "Publishing data in a reusable form to support findings must be mandatory."
2013	European Commission's Horizon 2020 research programme announced the Open Research Data Pilot, mandating DMPs and making underlying data available in a repository, but allowing those funded to opt-out. The Research Data Alliance is founded by European, American and Australian scientific government agencies, to build the social and technical bridges to enable data sharing and re-use.[9]	League of European Universities publishes LERU Roadmap for Research Data (Achard et al., 2013). Piwowar and Vision (2013), publish Data Reuse and the Open Data Citation Advantage, a highly cited paper giving evidence that data sharing improves citation rates. DCC publishes case studies of UK institutions whom they were funded to assist to develop RDM policy and services (Jones, Pryor, Whyte, 2013).

[7] https://sparcopen.org/our-work/innovator/panton/
[8] https://www.smartdatacollective.com/gartner-adds-big-data-its-2011-hype-cycle/
[9] https://www.rd-alliance.org/about-the-rda.

Year	Drivers/Milestones	Key Documents
2014	Results of the 2014 Research Excellence Framework announced, and a mandate for the next REF given: for the first time submissions of research outputs for quality assessment will be required to be open access.[10] LARD-LONDON Jiscmail list created for London area professionals working in RDM support to network.	FORCE11 publishes the Joint Declaration of Data Citation Principles (Martone, 2014).
2015	Engineering and Physical Sciences Research Council (EPSRC) publishes 'expectations' of institutional responsibilities for data curation and data management plans.[11]	Kristin Briney, 2016, a "chemist turned data librarian" publishes book, Data Management for Researchers.
2016	UK funders publish Concordat on Open Research Data, still current for what is now UK Research and Innovation (UKRI, 2016).	Desai, Ritchie and Welpton (2016) coin "The 5 Safes" as a framework for making sensitive data accessible for research. Wilkinson et al. (2016) publish the FAIR Guiding Principles, surpassing the OAIS Reference Model for most used slide in UK RDM presentations.

From about 2016 onwards, many UK institutions had hired at least one RDM professional and had settled on a service and policy approach. Jisc was no longer funding institutional development in this area, but the European Union through its successive Horizon funding programmes began investing in projects to encourage FAIR data use (Findable, Accessible, Interoperable, Reusable) by researchers, and solidified its support for both OpenAire and the Zenodo open access digital repository for free use by European researchers and beyond. The phrase from the Horizon 2020 DMP pilot, "as open as possible, as closed as necessary," became a well-known slogan, and data sharing was championed by senior academics such as Professor Geoffrey Boulton, in his role as CODATA President (2014–2018), and Professor Barend Mons, e.g. in his role as Chair of the High Level Expert Group of the European Commission on the European Open Science Cloud, and one of the originators of FAIR (Wilkinson et al., 2016).

This in turn led to a proliferation of discipline-specific efforts, with varying degrees of success, to bring FAIR data sharing and other forms of Open Science into the mainstream. Library RDM services began to find their true audience rather than their good practice messages being missed or ignored altogether; the central services could increasingly collaborate with equally enthusiastic researchers. Researchers' involvement has led to

[10] https://blogs.lse.ac.uk/impactofsocialsciences/2014/04/01/hefce-open-access-ref-gamechanger/
[11] https://ec.europa.eu/commission/presscorner/detail/en/ip_13_1257

interesting initiatives such as the Hidden REF[12], and the CRediT (Contributor Role Taxonomy)[13] citation movement, both meant to give credit — and, importantly, career rewards — to those who work with data on research projects and other "hidden" roles, rather than just principal investigators and lead authors on papers. The blooming of Open Research initiatives (Open Research generally being the preferred term to Open Science in the UK, in order to include the Arts and Humanities) has led to data sharing being considered one of about half a dozen 'pillars' of open research good practice.

Meanwhile, "Data stewardship," or a devolved group of part or full-time data stewards, began to be used in the Netherlands and other European countries as a way to ground good practice in disciplinary communities. In many cases these roles have been coordinated through institutional or national "competence centres," which aim to provide the expertise researchers need to practice open science/research. Ironically, the institutions with the most established central services, such as in the UK, have been slow to embrace the distributed data stewardship model, for reasons of additional staff costs, perhaps, and the reduction of grant funded collaborations with other European countries following the UK's exit from the EU.

Drilling Down to Institutional Level: Edinburgh and Oxford as Case Studies in Early RDM

Edinburgh

The University of Edinburgh (UoE) had a strong start in the early development of research data management services (2004–2016), with several relevant projects and services in place (Rice, 2022), including:

- A 5-star rated Informatics school leading often interdisciplinary data-driven research.
- Home of the National e-Science Centre (together with University of Glasgow), roughly 2002–2011.
- A merged Information Services (from circa 2003), made up of the Library and the Computing Service, with a central helpdesk.
- Home of EDINA national datacentre funded by Jisc, with a 'local' Data Library service providing data-related user support to university researchers.
- Home of the Digital Curation Centre since 2004, together with University of Glasgow, and previously UKOLN at University of Bath, STFC in Reading and the UoE School of Informatics.

Collaborations with the DCC and the DISC-UK group of data support and repository professionals at the Universities of Edinburgh, Oxford, and Southampton, and the London School of Economics and Political Science (LSE) led to early innovations, such as the Jisc-funded projects: the Data Audit Framework Pilot Studies (2008–2009); DISC-UK DataShare (2008–2010); Research Data Management Training (MANTRA, 2010–2011) and another online educational resource (as partner to University of North Carolina): the Research Data Management and Sharing MOOC (2016). The results of these projects fed into the RDM University Policy (2011, revised 2022), the RDM Programme (2012) and its later

[12] https://hidden-ref.org/about/
[13] https://credit.niso.org/

incarnation as the Research Data Service (2016). An academic-led steering committee has ensured 'fitness of purpose' and internal funding opportunities for the activities and service development over the years, in particular through successive Research Data Roadmaps, from 2012–2020.[14]

From 2018, the Data Library team from EDINA merged with the Research Data Support team of Library & University Collections (L&UC) and became a library-based team. Research Data Support acts as a "second-line" support team on the Information Services helpline, engages with researchers one to one on answering queries and supporting DMPs, leads on RDM training aimed at staff and research postgraduates, and operates the two data archiving solutions, DataShare and DataVault. The team works closely with the IT Infrastructure team which operates the institutional data storage solution—DataStore—and related applications (DataSync) plus high performance computing, and also with the Digital Library team which maintains and develops the two archiving solutions.

The team's support functions are amplified and promoted through a number of outlets, including Digital Research Services facilitators (a team of three college-facing specialists whose job is to raise awareness and facilitate access to all research-related digital services offered within the university); L&UC's Academic Support Librarians; Library Research Support, the section in the Library to which the Research Data Support team belong, and which owns the LERU-based University of Edinburgh Open Research Roadmap[15]; the Institute for Academic Development; Information Service's Digital Skills; the Research Outputs Network of college and school-based research support staff; the Edinburgh Open Research Initiative collective of staff and students; as well as a new group of institute-based Research Data Managers based in the College of Medicine and Veterinary Medicine.

Oxford

The University of Oxford has been offering research data management support services for several decades, with the Oxford Text Archive serving the needs of digital humanists as early as the 1970s. Over time different initiatives served different user needs; support has focussed mainly on helping researchers to understand and meet the requirements of different research funders, with a focus on practical data handling and good practice. In 2013, a 'Managing Research Data Programme' funded various projects, including DataStage, DataBank, and DataFinder, which served as the foundations of an institutional repository service.

Oxford's institutional data repository was launched in May 2015, in response to the Engineering and Physical Sciences Research Council (EPSRC) policy framework on research data. Prior to this, Oxford ratified its first Policy on the Management of Research Data and Records in July 2012—again, in response to an external driver: the Research Councils UK 'Common Principles on Data Policy', published in April 2011. The policy made commitments to University provisioning in specific areas, including 'services and facilities for the storage, backup, deposit and retention of research data and records', and 'access to training, support and advice in research data and records management'. ORA-Data was established to provide open access to research datasets, with metadata review, DOI assignment, and long-term storage and preservation. Engagement with the ORA-Data service has been consistent but fairly low. ORA-Data received its 100th deposit around one year after launch; to date, the service holds around 1,500 data records (around 1,300 with files available to download).

Provision for 'active' data is through IT Services (ITS), via a 'Research File Service' which has evolved over the course of a decade, providing secure and scaleable storage for 100GB–20TB of research data. Centralised provision at Oxford has been and remains challenging; Divisions and Departments have tended towards local solutions and staffing, and use of tools which are institutionally approved but not always appropriate for the task (OneDrive),

[14] https://information-services.ed.ac.uk/about/strategy-planning/rdm-roadmap
[15] https://library.ed.ac.uk/research-support/open-research/about

or not institutionally approved (Dropbox). ITS additionally provides a service for the secure storage and long-term archiving of data, including confidential data, called DigiSafe, which uses the LibNova platform. Finally, the Sustainable Digital Scholarship (SDS) service (previously called Digital Humanities Sustainability) is utilising Figshare, launched in 2021, to provide space for semi-active, open, project research data.[16], [17]

Provision at Oxford, then, is quite diffused and granular but has been developed based on user need as well as external factors; activities and service support are provided by a cross-University collaboration of colleagues from the Bodleian Libraries, IT Services and Research Services, named 'Research Data Oxford' and underpinned by a website and a virtual ticketing system. The Research Data Management Review in 2020 made a number of recommendations, including: strengthening university governance and policy; reviewing funding and embedding sustainability; extending and harmonising provision for active, semi-active and archive data; and developing and joining up training, services and support—this has shaped priorities and activities since, with the Bodleian undertaking further work in 2023 to develop its longer-term vision and actions in support of the Review. This included the project 'Consolidating the Bodleian Libraries' Research Data Management services' between September 2023–February 2024, during which a benchmarking exercise of research-intensive institutions was carried out.

Benchmarking Exercise: Key Findings

The aim of the benchmarking exercise was "To understand how the Bodleian's RDM services offer compares with those of peer UK institutions in terms of training, guidance and support, systems and infrastructure, and data services; and to inform the structure and roles of RDM support in Scholarly Resources and a roadmap to develop the Bodleian's RDM services." The full interview instrument (set of questions) is included in Appendix 2.

Generally, the RDM support team in a university library provides general RDM advice, guidance, and training, offers first line support and is responsible for the RDM website; specific domains are training and support for Data Management Plans (DMP) including DMPonline and deposit of research data into the institutional repository or an appropriate domain repository. Comparisons by topic for each institution follow.

Governance, Policy, Staffing

All the institutions were providing RDM services that followed an institutional RDM policy of some kind, which was regularly reviewed and recently updated (within the last three years). All of the institutions' services had formal governance that reported to official university committees, apart from Imperial College London—however, in this case, the Vice-Provost for Research's office intended to establish an institutional Open Research Group whose business would include RDM.

The library was generally found to be the place where RDM staff were employed. Although it is difficult to compare exact staff make-up, in many cases the library team was closely aligned with scholarly communications and open access publications/repository teams, and only responsible for certain aspects of RDM support. These dedicated RDM staff often worked closely with other centralised teams in computing: Edinburgh – Research Services in IT Infrastructure; Cambridge – Central University Information Services; Imperial – Information and Communication Technologies; Manchester – Research IT in IT Services; Oxford – Research Support in IT Services; UCL – Advanced Research Computing Centre.

[16] https://dl.acm.org/doi/10.1145/3627169
[17] https://www.tandfonline.com/doi/full/10.1080/14794713.2023.2206286

Unlike Edinburgh and Cambridge, front-line or policy support for RDM was often split between more than one central, Library-based team. These ranged from the research office (Imperial, UCL), to other Library teams (Oxford), to Research IT/Computing teams (Oxford, Manchester, UCL). Typically, library-based RDM teams deal with data management planning, sensitive data, and archiving requests. Edinburgh, in addition, handles active data queries as part of the central Research Data Service, with active data platforms mainly supported by Research Services (IT Infrastructure).

It appeared there were common challenges stemming from distribution of responsibilities not only between the centre and periphery, but amongst central teams based in different organisations—something which these research-intensive universities appear to have in common, although some (Edinburgh, Cambridge) have a more centralised home for RDM support, despite sharing the challenges of visibility across a large, distributed institution.

Long term preservation and storage

Perhaps unsurprisingly, given the nature and mission of libraries, long-term preservation and storage was a rich section of the interviews with the providers of open access data repositories as a core aspect of service provision. Each institution hosts some form of open access data repository, whether built in-house on open-source software (generally DSpace or Fedora) or by cloud subscription (Figshare). Imperial were looking into InvenioRDM for a new in-house repository, an open-source solution used by Zenodo and others. Only two stated that they have the CoreTrustSeal trustworthy digital repository certification for their repository—Edinburgh DataShare, an institutional multidisciplinary research data repository, and Apollo at Cambridge, which also stores publications and other research outputs—with two more, Oxford and UCL, intending to start the application process. Most mentioned directing users to discipline-specific repositories first. Imperial also directs users to the generalist repositories Zenodo, Figshare and Dryad. Edinburgh's DataVault was the exception, designed for sensitive data deposits as well as large data. However, all identified sensitive data and large data as difficult requirements to resolve, and some were attempting to address sensitive data through their Trusted Research Environments (TREs) rather than repositories. Deposits range from low up to 30-50 deposits per month, but some accept ephemera, not just research data.

UCL has a Figshare repository currently. Manchester also uses Figshare in a way that interoperates with Pure, the research information system or CRIS (Current/Campus Research Information System). The Manchester RDM service team review each deposit. A deposit limit of 20 GB can be overcome through a File Transfer Protocol (FTP) interface. UCL has a limit of 50 GB for individuals or 100 GB for group projects, but increases are currently granted without additional charge. Edinburgh DataShare is free with up to 100 GB deposits and DataVault charging upwards of 100 GB with the highest deposit limit by far, at 10 TB. DataVault serves two purposes: for retention of data too large for the open access repository, and for retention of data too sensitive for the OA repository.

Oxford Research Archive has invested in years of digital preservation development and has predictive file storage usage and growth capability. Development in the near future will increase the deposit limit to 256 GB and implement DOI versioning.

The benchmarking exercise covered whether each institution had a data catalogue, implicitly defined as records about datasets held in other repositories. Most used either their repository or their CRIS as a data catalogue (Cambridge did not have one) and most had plans for harvesting information automatically to populate this information. This is an area which is likely to expand and solidify in future. Edinburgh uses Pure for its data catalogue, which also acts as the University's CRIS, an essential tool for the REF.

Trusted Research Environments (TRE)

Although not an explicit benchmarking category, TREs emerged as a clear growth area and a potential solution to the tricky question of sensitive data (with Edinburgh the only institution, at the time of interview, which took sensitive data via its DataVault service).

Cambridge, Manchester and Oxford each have an Economic and Social Research Council (ESRC)-funded SafePod administered by the Library. Edinburgh has not been able to find a suitable setting in the Library for a SafePod (due to pressure for increased student working space) but is seeking an alternative location for ESRC safe points (specially equipped desks) in partnership with a school.

Edinburgh TRE provision was the most centralised and mature, with the International Data Facility, run by Edinburgh Parallel Computing Centre, hosting a number of TREs including an NHS Lothian Data Safe Haven, a managed 'Dataloch' of highly consulted health datasets, and bespoke environments with high performance computing capability, though this is separate from the Research Data Service, and not all research projects are necessarily eligible.

Provisions at other HEIs was more localised and bespoke, or in the pilot phase: Cambridge (Department for Psychiatry) has a "RedCap Safe Haven service", which is a chargeable service offered to other university users in which direct connections with Apollo are currently being explored. UCL provides ISO 27001 certified TREs for its medical school, whilst Imperial has started a pilot to provide the same.

Oxford is in the planning and development stages for a TRE Framework for the University, as part of the Digital Transformation project.

RDM Tools and 'Data Services'

Data tools and services was another area of mixed provision and potential growth. All institutions provide access to databases via library catalogue search and guides, and all used DMPOnline, though Cambridge has not yet taken a subscription, referring users to the free version.

For Text and Data Mining (TDM) activities, support varied from provision of online guidance, to tools, training and queries, with Oxford and Manchester subscribing to ProQuest TDM Studio, and Oxford to Gale Digital Scholar Lab. Oxford has a dedicated 'Data Access Librarian' role which provides direct support with access to, and use of, specialised (usually charged-for) data resources.

In terms of specialist tools, Oxford and Edinburgh subscribe to the Open Science Framework, with Edinburgh hosting a GitLab instance and using Globus for large file transfers. At the time of interview, there was little to report on AI tools other than guidance such as LibGuides, although all institutions expressed an interest in monitoring use and provision of such tools, and in the potential of AI in relation to RDM services.

Communication, training and events

Activities in this area were remarkably similar, including the content of regular training sessions (RDM practice, data management planning, and sensitive data). UCL and Imperial offer embedded training for DPhil students, and Manchester offers core RDM training via an centralised research programme.

Edinburgh also has a data archiving course as a regularly scheduled session, a self-training online course (MANTRA), and runs an annual Open Research Conference (formerly Dealing with Data). Cambridge, Oxford and UCL use Love Data Week, an international event falling over Valentine's Day, to showcase support for RDM.

Cambridge also separates training content for STEM (Science, Technology, Engineering and Mathematics) and SHAPE (Social Sciences, Humanities and Arts for People and the Economy/Environment) audiences. In the case of sensitive data, Oxford offers separate

content for social sciences and medical audiences. All of the institutions utilise a ticketing system for user enquiries with the exception of Manchester, which handles email enquiries via a central inbox and assignment across the RDM team.

Data champions and data stewards

This was another identified growth area, with variable provision and much interest from participants. Cambridge has a mature programme with 180 people identified as Data Champions across the University. The programme started in 2016. RDM support staff spend 20% of their time tending this network. UCL has a bank of 12 data stewards who can be assigned to research projects according to skillset and are costed into research grants, similar to Research Software Engineers (RSEs). Edinburgh has data managers hired by individual projects and more recently, by some of the Medical School's institutes, and Digital Research Service Ambassadors (postgraduate students) who are paid to work with research projects to add a digital skillset. Manchester was about to begin a large-scale data stewardship project, starting with Data Steward coordination role in the Library. Everyone has identified the need for research data stewards or champions to help distribute the load and disseminate key messages across the breadth of the institution.

Digital Humanities

For those institutions with Digital Humanities provision, the university libraries work closely with academic digital humanities initiatives to support events, training and research, through support of Digital Humanities Labs (Cambridge) or Centres (Manchester), Centre for Data, Culture and Society (Edinburgh); Oxford has a Centre for Digital Scholarship, which started as Digital Humanities and later expanded.

For Humanities-specific data storage: several institutions support legacy or new 'warm' (semi-active) project data; Oxford and Manchester both use Figshare to store 'warm' or semi-active data in collections. Cambridge and Edinburgh support warm data through existing RDM functions, such as the IIIF digital viewer (Cambridge). Neither UCL nor Imperial systematically support warm data, with UCL hosting bespoke platforms and Imperial focusing on hot and cold data as befits their more scientific remit.

Future directions

These are the themes the authors feel the UK institutions have in common regarding future directions.

- Artificial intelligence and other tools
- New horizons for TREs and information governance
- The centre and the periphery, and nascent data steward networks
- Those who champion or steward data
- The growth of RDM situated in open research policy

A brief summary of each theme follows, along with a summary of future plans for Edinburgh and Oxford.

Artificial intelligence and RDM

At the time of the interviews, most institutions were still grappling with use of generative AI as an academic tool, and no bespoke use for RDM purposes were reported. However, as institutions create protected platforms for student and academic use, such as the Edinburgh Language Model (ELM), options may begin to emerge, for example: chatbots for IT, Library and RDM local services automating some aspects of assisting users; pre-filled metadata field suggestions for repository deposit based on uploaded files; use of Microsoft Copilot or similar by researchers and Research Software Engineers to write scripts for data handling/wrangling and to improve or speed production of research software code (this is undoubtedly already happening, if not as a service).

New horizons for TREs/Information Governance

Just as TDM services for Digital Humanities scholarship have evolved largely through partnerships between the Library and academic (SHAPE) departments, it is also the case that those institutions going beyond research funder-supplied Safepod safe settings to offering actual trusted research (server-based) environments, have largely been able to do this through partnerships with academic departments (e.g. Cambridge's Redcap service and UCL's TREs based in the Medical School).

Information governance, like Ethics, is largely considered to be a function served within academic departments too. But it is difficult to do RDM and data stewardship well without tackling information governance. We believe that libraries, as they hold ageing research data assets, will need to form alliances with university legal services (who provide contracts, MOUs and data sharing agreements for research projects), and possibly even form formal data access committees, such as is done at Bristol University (outside the interviewed group) to make decisions about data requests of institution-held data, and their ultimate removal.

The Centre and the periphery, and nascent data steward networks

Large institutions such as those interviewed for Oxford's benchmarking study are familiar with tensions between central services, such as university-wide IT services, and local providers based in academic departments. The glimpses we have into the workings of RDM at these different institutions show not only how centralised departments such as Library, IT and Research Offices interoperate to provide the spectrum of RDM services, but also the dual challenge of spreading knowledge of RDM good practice to those who support academics and research students more intimately (training the trainers), while keeping control of the services and key messages. Looked at the other way, those providing local support or signposting with knowledge of disciplinary behaviour may be frustrated by the necessarily generic aspects of the central services and the lack of customised RDM and open research messages that could be more tightly focussed on 'their' researchers and types of research done.

The data champions network that has been 'championed' by Cambridge, or the ReproducibiliTea student reading groups started at Oxford, or the grouping of Research Software Engineers for hire in some institutions are effective examples of ways the centre and periphery can interwork effectively. There is some indication this tendency may develop in future through networks of data stewards (be they novices eager to learn, or paid staff who can be booked onto projects, as with UCL). There is a European model, funded and encouraged through current European Commission projects, which federates data stewards either at a single institution or nationally through Digital Competence Centres. Whether current strains on UK university funding will restrict the development of such networks in the UK is yet to be seen.

The growth of RDM situated in open research policy

A final observation on where we have come from and where we are heading is in the area of university policy. RDM policies emerged in the UK largely as a response to UK funders' requirements placed on researchers and their institutions, but they were largely grounded in the context of research integrity (or the opposite of scientific misconduct). Future policies will likely have even stronger contexts within open research (or science), research culture, and the rewards and incentives springing from those initiatives. FAIR has given us the nuances needed to address skepticism about 'openness', e.g. data need not be purely 'open access' to be considered discoverable and accessible. RDM feels like a pillar of the broader open research movement increasingly, and policies will increasingly reflect this natural 'home'.

Additionally, the RDM policies of UK universities will undoubtedly need to shore up the data polic(ies) of UK Research and Innovation (UKRI), the primary funder of UK-based research, made up of the discipline-specific research councils. UKRI's 2025 exercise of updating and creating an over-arching UKRI data policy will undoubtedly affect future university research data management policy updates.

Oxford plans

Research Data Management remains a strategic priority across the University, with an immediate focus is on improving the visibility and join-up of services so that researchers can find information and tools relevant to their needs, when they need them. The Bodleian Libraries and IT Services are working together to improve and formalise the infrastructure behind Research Data Oxford and to ensure that University storage provision is harmonised and linked logically. Additionally, the Bodleian Libraries is taking forward its plans to consolidate and expand its services, focussing on: building on a comprehensive 2023 review of repositories; expanding adjacent services around DOIs and open infrastructure; increasing core training and developing the knowledge and skills of subject librarians, as they are embedded in academic departments. Additionally, the Bodleian has set up an AI Task & Finish Group to address growing concerns from researchers in the arena of open research, —including data, and reuse by LLMs—and is seeking to develop new policy in this area, which will inform future service provision.

Edinburgh plans

The University of Edinburgh's Research Data Service, by now, is a relatively mature service, considered stable and useful to the research community, while continuing to increase take-up and expand its portfolio according to demonstrated need. It is being developed in two specific directions:

1. That of Open Research, led by the Library, with individual and groups of academic champions organising activities, and aligned with university commitments such as the Research Cultures Action Plan.[18] By being grounded within Open Research, the service actively steers users toward more open options for making scientific and scholarly research available, as well as making the systems that support the REF, Open Access publications, data and code more interoperable.

2. That of the Digital Research Services[19], a programme of outreach and activity to help researchers access all of the data and computing services available at the university, organised through a research lifecycle framework. By being grounded within Digital

[18] https://www.ed.ac.uk/research-innovation/research-cultures/research-cultures-action-plan.
[19] https://digitalresearchservices.ed.ac.uk/

Research Services, the service can be developed through investment aimed at increasing the capacity of university researchers to competently be able to use the most appropriate digital technologies in their fields and in interdisciplinary collaborations, with no gaps in their access to safe storage, analysis and compute environments, including new and innovative methods involving data science and artificial intelligence.

Acknowledgements

The original benchmarking interviews took place in January 2024 and were conducted by Lotte Boon, project lead; Ruth Mallalieu, Head of Open Scholarship Support; John Southall, Data Librarian; Jason Partridge, Open Access Service Manager; and David Tomkins, Research Data Curation Specialist.

The authors wish to thank the following library colleagues, who participated in the interviews and fed back on the benchmarking profiles, and who have advised on the development of this paper:

- Clair Castle, Research Data Manager, University of Cambridge
- Wayne Peters, Research Data Manager, Imperial College London
- Bill Ayres, RDM Strategic Lead and Clare Liggins, RDM Coordinator, University of Manchester
- Kirsty Wallis, Head of Research Liaison, University College London (UCL)

And, finally, special thanks and acknowledgement to Lotte Boon, who organised the interviews, and who compiled the original benchmarking data on which this paper is based.

References

Achard, P., Ayris, P., Fdida, S., Gradmann, S., Horstmann, W., Labastida, I., Smit, A. (2013). *LERU Roadmap for Research Data: LERU Advice Paper*. League of European Research Universities. Retrieved from https://www.leru.org/publications/leru-roadmap-for-research-data.

Briney, K. (2015) Data Management for Researchers, Pelagic.

Carr, D. (2017). Developing and Implementing the Wellcome Trust's Data Management and Sharing Policy. In: *LEARN Toolkit of Best Practice for Research Data Management.* (pp. 10-13). doi: 10.14324/000.learn.02.

Chiarelli, A., Beagrie N., Boon, L., Mallalieu, R., Johnson, R., Warner May, A., Wilson, R. (2022). To protect and to serve: developing a road map for research data management services. *UKSG Insights*. Retrieved from https://insights.uksg.org/articles/10.1629/uksg.566

Consultative Committee for Space Data Systems (2012). *Reference Model for an Open Archival Information System (OAIS), Recommended Practice, Issue 2, International Standards Organisation*. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284.

Desai, T., Ritchie, F., Welpton, R. (2016). *Five Safes: Designing Data Access for Research. Working Paper.* University of the West of England. Retrieved from http://eprints.uwe.ac.uk/28124.

Gooch, M., Strange, D. (2023). Consolidating Research Data Management Infrastructure: Towards Sustainable Digital Scholarship. *ACM Journal on Computing and Cultural Heritage.* doi: https://dl.acm.org/doi/10.1145/3627169.

Heery, R., Anderson, S. (2005). *Digital Repositories Review*, University of Bath Online Publication Store. Retrieved from http://opus.bath.ac.uk/23566/2/digital-repositories-review-2005.pdf.

Jones, S., Pryor, G. & Whyte, A. (2013). How to Develop Research Data Management Services - a guide for HEIs. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Retrieved from https://www.dcc.ac.uk/guidance/how-guides/how-develop-rdm-services.

Lord, P., Macdonald, A, Lyon, L., Giaretta, D. (2004). *From Data Deluge to Data Curation*. Retrieved from https://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf.

Martone, M. (ed.) (2014) *Joint Declaration of Data Citation Principles, Data Citation Synthesis Group*. FORCE11: the Future of Research Communications and e-Scholarship. Retrieved from http://www.force11.org/group/joint-declaration-data-citation-principles-final.

OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris, https://doi.org/10.1787/9789264034020-en-fr.

Open Research Data Task Force (2018). *Realising the Potential: Final Report of the Open Research Data Task Force.* Warwick: England. Retrieved from https://assets.publishing.service.gov.uk/media/5c51bf9a40f0b6254b1b1035/Realising-the-potential-ORDTF-July-2018.pdf.

Piwowar, H. A. and Vision, T. J. (2013). Data Reuse and the Open Data Citation Advantage. *PeerJ*, 1, e175. http://doi.org/10.7717/peerj.175.

Rice, R. (2022). A Decade of Research Data Management at the University of Edinburgh: Looking Back, Looking Forward. In Mani, N. & Cawley, M. (eds.). *Handbook of Research on Academic Libraries as Partners in Data Science Ecosystems*. Hershey, Pennsylvania, USA: IGI Global. p. 308-333. doi: 10.4018/978-1-7998-9702-6.ch015. Retrieved from https://www.research.ed.ac.uk/en/publications/a-decade-of-research-data-management-at-the-university-of-edinbur.

Strange, D., Gooch, M., Collinson, A. (2022). Equality, findability, sustainability: the challenges and rewards of open digital humanities data. *International Journal of Performance Arts and Digital Media.* Retrieved from https://www.tandfonline.com/doi/citedby/10.1080/14794713.2023.2206286.

The Royal Society Science Policy Centre (2012). *Science as an Open Enterprise.* London. Retrieved from https://royalsociety.org/news-resources/projects/science-public-enterprise/report/.

UKRI, (2016). *The Concordat on Open Research Data*. Retrieved from https://www.ukri.org/manage-your-award/good-research-resource-hub/open-research/.

Eynden, V., Corti, L., Woollard, M., Bishop, L. and Horton, L. (2011). *Managing and Sharing Data; a Best Practice Guide for Researchers.* Essex, England. Retrieved from https://dam.ukdataservice.ac.uk/media/622417/managingsharing.pdf.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data.* doi: 10.1038/sdata.2016.18

Appendix 1: Matrix of interview responses (summary)

	Cambridge	Oxford	Edinburgh	UCL	Imperial	Manchester
Governance and policy	RDM Policy Framework owned by Head of Open Research Services, Office of Scholarly Communication (OSC), University Library. RDM is responsibility of Open Research Steering Committee, which reports to Research Policy Committee.	RDM Policy is the responsibility of the Research Data Steering Group. The policy is amended by the RDM Delivery Group, which consists of members of Research Data Oxford (Bodleian, IT Services, and Divisional reps) and requires sign off by Research and Innovation Committee, which is chaired by the Pro-Vice Chancellor for Research.	RDM Policy owned by Library Research Support. RDM is responsibility of Research Strategy Group, while Research Data Steering Group has most immediate responsibility. Digital Research Services Steering Group provides internal funding for development.	Research Data Policy owned by Advanced Research Computing Centre and reviewed by the Research Data Working Group, which is convened by Library, Culture, Collections and Open Science (LCCOS). Responsibility for research data strategy has been transferred to LCCOS, alongside other strategies in the related areas of research practice.	RDM Policy is owned by the Research Office. No formal governance of RDM service. The Vice-Provost for Research's office intends to establish an institutional Open Research Group—reporting into the University Research and Enterprise Board—which will be co-led by a senior academic and the Director of Library Services or Research Office.	RDM Policy is owned by University Library. RDM is the responsibility of the Library Executive Team. RDM Operational Group reviews minor policy changes. The RDM Standard Operating Procedure document is reviewed at the same time as the policy; they are linked.
Support team	Based in OSC, University Library: c. 8 roles with research data responsibility and alignment. Research	No centralised team, roles spread across library, c. 9 with research data responsibility and alignment.	The Research Data Service (RDS) team consists of c. 6 staff with direct RDM responsibility: Data Librarian & Head,	RDM support is mainly provided by LCCOS, with c. 3 staff with direct RDM responsibility. Head of Research Liaison (OS &	The RDM team in Library Services serves as the central hub for RDM. Small team of c. 3 staff with research data	The Research Data Management Service (RDMS) team is based in the Office for Open Research in the library, c. 6 staff with

	Data Manager, 2 x Research Data Coordinators, Head of Open Research Systems (OA & RDM), Repository Manager (OA & RDM); Scholarly Communication Assistant (OA & RDM), Scholarly Communication Support role (OA & RDM), Data Analyst (OA & RDM).	Head of RDM, Research Data Curation Specialist, Open Access Services Manager, OA Supervisor, Data Librarian (Bodleian); 4 x SDS team (Bodleian & Humanities); 5 x Research Support Team (IT Services).	Research Data Support (strategy and stakeholders); Research Data Support Manager (operations); Open Research Coordinator / Research Data Support Officer; Research Data Support Officer (outreach, training, DMP); 2 x Research Data Support Assistant (1.5 FTE DataShare and DataVault).	RDM); 2 x Research Data Support Officers. Advanced Research Computing (ARC) provides RDM support for infrastructure and technical advice (re. the Research Data Storage Service and the UCL Research Data Repository).	responsibility and alignment. comprising: Open Research Manager: Academic Engagement; Research Data Manager; Senior Scholarly Communications Assistant (RDM & Open Research.	direct RDM responsibility: RDM Strategic Lead; RDM Coordinator; 2 x RDM Librarians; RDM Analyst; Analyst and Support Assistant.
Preservation and storage	Apollo uses DSpace. Apollo holds a CoreTrustSeal. Elements deposits up to 2GB), can be compressed and deposited. One-off charge for long-term curated data storage of £4/GB for datasets above 20GB. 1-2 deposits received per day (~28 a month). Does not take sensitive data. Does not have data	ORA uses Fedora / Hyrax. Applying for CoreTrustSeal. Elements deposits up to 2GB; a large file upload (up to 256GB) about to be delivered. No charge for storage. Around ~15 deposits per month. Does not take sensitive data. New digital preservation system using Fedora 6 (OCFL), provides robust back-up and restore mechanism.	DataShare uses DSpace. DataShare has the CoreTrustSeal. Takes data up to 100GB. No charge for direct deposits. Larger files go to DataVault – a separate charged service. Around 20 deposits per month. Does not accept sensitive data. Pure is the University's data catalogue. A metadata record in Pure is automatically created for deposits into	UCL repository uses Figshare, applying for CoreTrustSeal. Files larger than 5GB can often not be uploaded using the browser—use Figshare FTP uploader for large files. Users receive a storage limit of 50GB or 100GB for group projects but can request storage increases without extra charge. About ~15 deposits per month. Users must create a metadata record of	Imperial has a bespoke, in-house data repository (InvenioRDM being considered). Maximum file size limit of 1GB, does not accept sensitive data. Symplectic Elements and the repository for publications, Spiral, serve as data catalogue. Exploring data catalogue from open datasets in other repositories using Scholex API.	Uses FigShare. Data is pulled through to Pure and records can be accessed in Research Explorer, the public portal of Pure. Single deposits no larger than 20GB, FTP uploader available for larger datasets. No limit to the size of dataset. ~40 deposits per month. Sensitive data is not taken. Pure is the University's data catalogue. Users are asked to use the

	catalogue. DOIs are registered for content. Looking at Fedora 6 for digital preservation solution.	Microservices also run on ORA and systems. Looking to use ORA as a data catalogue (not yet established). DOIs are registered for content.	DataShare and DataVault. Using Pure Data Monitor to harvest information about datasets located outside Figshare. DOIs are registered for content.	datasets in the UCL Research Publications Services (RPS) or UCL Research Data Repository. Sensitive data is not taken.	DOIs for items in the data repository, theses and grey literature.	Research Data Gateway if their dataset has a DOI for a record to be created in Pure. Using Pure Data Monitor to harvest information about datasets located outside FigShare.
Training, advice, support	Research Data website. Queries use ticketing system. Library-based skills training. No Subject Librarian support - RDM training delivered by Research Data Coordinators; Research Support Librarians contribute. Bespoke subject area training. Free version of DMPonline; DMP review available. Data Champions programme since 2016, ~185 as link between libraries and departments.	Research Data Oxford (RDO) website. Library info iSkills website and IT learning website. Uses ticketing system and consultation. Library and IT RDM training general and bespoke. Subscription to DMPonline. Review service for DMPs. Additional training incl. library induction, MSc in Digital Scholarship & Digital Humanities at Oxford Summer School. Subject Librarian support. Bodleian 'data leads' link between Libraries and Divisions.	Information Services and Digital Research Services websites. 4 training workshops and 3 online courses. Targeted training for new researchers (Institute for Academic Development). UoE specific DMPonline subscription; queries submitted to RDS through DMPonline. Subject Librarian support. Digital Research Services have paid student ambassadors (Research Software Engineers) to help with coding, etc. on research projects.	Website section for RDM support. Regular online and in person training for staff and students. RDM part of PhD training. Training advertised as part of library skills; on social media; and online. DMP training and advice provided. 12 data stewards can be costed to grants for research projects. May introduce data champions as sensitive data service develops.	General advice provided by Scholarly Communication (website section). RDM website. RDM training termly for Graduate School; postdocs; and fellows. DMPonline subscription. DMP review. No Subject Librarian support. Research Software Community has research software champions; scoping creation of data champions or stewards.	RDM support via institutional and Digital Research Services websites. Workshops and online courses available on basic RDM. Subscription to DMPonline; DMP review available through this system. Research Services Librarian and the Engagement Librarian support provided. Implementing data steward roles, starting with Data Steward coordination role in Library.

		Scoping data champions or stewards.				
Data services	Access to databases provided via the library search. Parallel services give access to clinical data and institutionally created datasets (Cam:ID and Apollo). A service to give access to trusted research environments for restricted data is in place (using an ESRC-funded Safepod managed by the library). Cambridge Integrated Data Environment (CAM:IDE) offers RedCap Safe Haven service.	Access to databases is provided via the library; this includes a dedicated Data Librarian. Parallel services give access to institutionally created datasets (ORA and SDS). A service to give access to trusted research environments for restricted data is in place (using an ESRC-funded Safepod managed by the library). The Bodleian Data Service team provides additional support in access and training around using research data. Institutional TRE framework project due to start.	A list of key data resources is provided on the university website. Datasets are also searchable on the main library guides and search engines. A service is in place to give access to ONS restricted data (only for some of the schools; not the whole university). Scoping alternative location for ESRC safe points.	Data services are mainly confined to traditional signposting of datasets via library guides and search engines. TRE – Data Safe Haven. There are also a few trusted research environments within departments for accessing specific restricted datasets.	Not covered in this particular interview. TRE pilot underway for sensitive medical data.	Library guides and the main search engine include datasets. Restricted data access is supported by use of an ESRC-funded Safepod. Additional support on creating and working with restricted data is provided by the 'Highly Restricted Data Service' in development (managed by library and IT services). An increase in the number of open research platforms supported and promoted by the library is also being pursued.
Tools	Library offers TDM and AI tools guidance through LibGuides and	Library advises on TDM for bibliographic resources, but licensing doesn't allow	Digital Library support: DMP online, OSF, GitLab, DataSync (file sharing service),	TDM Libguide lists APIs by publisher and directs researchers to APIs from sources e.g.	RDM team support DMPOnline only; include open source tools e.g. Openrefine	Library provides expert support incl. use of TDM, and subscribes to Proquest

	additional support via the Ejournals team. They do not pay for TDM with subscription databases. OSC interested in use of AI in scholarly communications e.g. DMP creation and publishing.	for TDM use with other subscription collections. Subscribes to OSF. University research funding given to understand potential for wider TDM, AI support. Since interviews TDM subscriptions in place for HathiTrust, Gale Digital Scholar and Proquest TDM Unlimited.	and Globus DataStore (secure high-performance data transfers from DataStore to external organisations & individual users), TDM and AI.	CrossRef, Scopus, or WoS. E-resources offer support where content providers don't have APIs. LCCOS doesn't fund TDM (subscription databases) with library budgets—researchers self-fund. LCCOS have an AI working group. Get enquires about specific AI tools but this isn't currently embedded in library training.	in their training. TDM queries handled via copyright specialist. RS Community will look at AI (from academic perspective/ funded by Research England).	TDM Studio (highlighted as a resource for Digital Humanities researchers).
Digital Humanities	Library supports academic Cambridge Digital Humanities (CDH) through DH Lab and in CDH learning programme. Digital Library (CUDL) hosts warm/semi-active data projects.	Library supports academic Digital Scholarship @ Oxford. Warm/semi-active data through Sustainable Digital Scholarship (Figshare) use of collections feature.	Library supports academic Centre for Data, Culture & Society which has some data support. Warm/semi-active data supported through existing functions.	Academic Centre for Digital Humanities is more a forum than research centre. No centralised support for warm/semi-active data; use of project websites.	No humanities. The RDM team support mainly at start (hot/ active) and end (cold/ archived) of projects.	Library supports academic Centre for Digital Humanities. Warm/semi-active data stored using Figshare collections and sub-groups.

Appendix 2: Interview instrument

Aim

To understand how the Bodleian's RDM services offer compares with those of peer UK institutions in terms of training, guidance and support, systems and infrastructure, and data services; and to inform the structure and roles of RDM support in Scholarly Resources and a roadmap to develop the Bodleian's RDM services.

Benchmarking questions

Research Data Management governance	
Do you have an RDM policy?	
Who owns the policy and how often is it reviewed?	
Which committee(s) or group(s) are responsible for RDM	
What do you think is good about your set-up and what could be improved	

Research Data Management support team	
Could you describe how research data management support is organised in your library. What are the roles, where are they based, and what is the FTE?	
Could you describe your RDM services?	
Do your research services department and IT services also support Research Data Management Could you describe this support? How well do you work with each other? What structures do you use? Are you based in the same building?	
Are you expecting any changes in this set-up?	
What do you think is good about your set-up and what could be improved	

Long term preservation and storage	
What is the name of your institutional repository	
Can researchers deposit research data in your institutional repository	
If yes, direct or via your research information system?	
Is there a limit to the size of the dataset	
Roughly, how many deposits do you receive per month?	
Do you accept sensitive data?	

How do you market your repository – is it repository of last resort, i.e. when there is no suitable subject repository?	
Is your repository also a data catalogue	
If yes how do you harvest records of data deposited elsewhere?	
Do you mint DataCite DOIs	
What is your data architecture, do you have DOI versioning and viewing by item/record level?	
Do you make use of Fabrica's functionality of having multiple repositories?	
What are your future plans?	

Data services The Bodleian Libraries provides a range of services for researchers and students at the University of Oxford who need to make secondary use of statistics and data – primarily focused on social sciences	
Do you provide a data service in the library? If yes, does it comprise a full list of data resources: commercial subscriptions, institutional repositories and specialist archives, restricted data repositories and archives, Open data portals Are these lists tailored in any way, e.g. by discipline	
Do you manage, support and advice on access to restricted data repositories incl. Eurostat, the Secure Data Lab of the UK Data Service and ADR UK?	
Our Social Sciences library has a SafePod for data that requires secure access for research purposes. Do you provide a similar service, a trusted research environment or are you working towards this?	
Do you provide data services we have no yet covered?	
What are your future plans?	

Training, advice and support	
Do you have a dedicated website for RDM support?	
What type of training do you provide and how is it advertised	
Do you subscribe to DMP online? Do you provide training in this area?	
Do you have a data steward role, if yes, how many data stewards do you have, where are they based and how are they supported, and how well does it work?	

Do you have a data champion role, if yes, how many data champions do you have, where are they based and how are they supported, and how well does it work?	
Do you expect your subject librarians to have a sign posting role re RDM? How are they trained and supported?	
Do you have an advisory service, what is the make up of the team who respond?	
Do you organise events: drop in events, roadshow, show and tell?	
What are your methods and channels of communication – how successful are these?	
How well do you think you're doing in this area and are there plans for change/improvement	

Research Data Management tools	
How do you support TDM with library collections, including working with data suppliers and implementing the TDM copyright exception	
In terms of TDM with subscription databases, which APIs for programmatic access to the raw data and TDM browser-based platforms do you support and how are these funded?	
Do you have plans or have you started looking into AI tools for bibliographic research, and embed within library training	
Do you provide other RDM tools?	
What are your future plans?	

Digital Humanities Within Oxford, the libraries will be taking on from our Humanities Division, support for digital humanities projects using Figshare as the platform and supporting digital humanities projects which revolve around a collection that is being added to periodically, and is a resource for other researchers	
Do you support semi-active research data, i.e. data published at the record / object level, that can be added to, as new records are created / discovered	
If yes, how is it funded and supported?	
If not do you have plans to develop support for semi-active research data?	