



Directions for Research Data Management in UK Universities

March 2015

Authors

Sheridan Brown, Rachel Bruce and David Kernohan

In collaboration with:



Contents

| | |
|---|-----------|
| Introduction | 4 |
| Vision | 6 |
| Five key areas for action: | 7 |
| Policy development and implementation | 7 |
| Skills and capabilities | 11 |
| Infrastructure and interoperability | 14 |
| Incentives for researchers and support stakeholders | 18 |
| Business case and sustainability | 21 |
| Moving forwards | 25 |

“Directions for Research Data Management
in UK Universities”

Authors
Sheridan Brown,
Rachel Bruce and David Kernohan



© Jisc
Published under the CC BY 4.0 licence
creativecommons.org/licenses/by/4.0/

Introduction

The growth of collaborative research practices in a connected era, and latterly a rise in funder mandates, have fuelled a rapid increase in interest in sharing research data.

Research data¹ is recognised as being central to research and disseminating, sharing and enabling access to research data are all now seen as essential to research integrity. Making research data accessible does not simply facilitate validation, it also supports new research and innovation.

Digital technology is making data sharing much easier.

Starting out

Openness is increasingly accepted as the default, while giving due respect to data protection, privacy and reasonable rights to first use. And so, as the highly regarded Royal Society report “Science as an Open Enterprise” says:

“The conduct and communication of science needs to adapt to this new era of information technology.”²

The opening up of research has also been supported by the transition to Open Access for peer reviewed research papers, and funders, publishers and universities have worked together to achieve Open Access. Open Access policies³ also encourage statements on access to the underlying research materials such as data. In the UK and further afield great strides have already been made to start to ensure that research data is well managed. The Australian National Data Service has coordinated developments for data infrastructure⁴ and, in the UK, Jisc has worked with numerous universities to develop solutions.

But the management of research data is not solely the concern of universities; research funders also have a role and around the world data management plans and access to data beyond the research grant period are now commonly mandated. In the UK research councils have

led the way with explicit policy requirements on research data and - alongside institutionally focused solutions for research data - there are disciplinary arrangements that cater for some data needs, such as the European Bioinformatics Institute⁵ and some of the research councils have disciplinary data archives.

Though the majority of research data is now in the digital domain, data does not have to be digital - the EPSRC⁶, for example, requires that “publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received”. Whilst this document focuses on digital data, much of what is said is applicable to all data.

Universities that have begun to address research data management actively have found that they need a multidisciplinary team - people in the research office, the library and the IT department may all need to find effective ways to pool their skills. Progress is being made but research data management solutions in UK universities are still relatively immature and most universities only have partial solutions in place. Many would welcome shared solutions, and it is clear that shared services and more tightly coordinated infrastructure is required for more efficient and effective research data management.

These are some of the motivations that led the Association of Research Managers and Administrators (ARMA)⁷, Jisc, Research Libraries UK (RLUK)⁸, the Russell Universities Group of IT Directors (RUGIT)⁹, the Society of College, National and University Libraries (SCONUL)¹⁰ and the Universities and Colleges Information Systems

Association (UCISA)¹¹ to discuss a joint direction of travel for universities over the crucial next five years. At the same time, Jisc has embarked upon a related area of research and development, namely Research at Risk¹², which focuses on finding and developing solutions for research data management within universities. Research at Risk seeks to realise a robust and sustainable research data management infrastructure and services to enrich UK research.

This report addresses five key topics:

- » Policy development and implementation
- » Skills and capability
- » Infrastructure and interoperability
- » Incentives for researchers and support stakeholders
- » Business case and sustainability

For each topic we have included a summary of the main current issues, alongside a vision of where the sector should aim to be in five years’ time. We then suggest actions for each topic, divided into ‘first steps’ and then longer term, more complex priorities. Readers should note that each of the five topics do raise interrelated actions, for example, a usage statistics service is flagged as a potential infrastructure solution and this issue arises again as an action area that can help to incentivise research data management and sharing.

Though we draw on selected recent publications, some stakeholder interviews and the outcomes of the Jisc’ Research at Risk’ consultation, much of the material here comes from a two day workshop held in Cambridge during November 2014, which was supported by ARMA, Jisc, RLUK, RUGIT, SCONUL and UCISA. Resources from the event are available online¹³.



Research at Risk

The consultation and development process leading to this report has been supported by the Research at Risk co-design challenge led by Jisc in partnership with RLUK, RUGIT, SCONUL and UCISA, and informed by numerous stakeholder consultation events. Indeed, one of the early ideas to come out of the consultation events was for a sector-owned direction of travel for RDM. This report represents a significant contribution to that. Many of the planned Research at Risk work areas directly address the needs raised within the report.



- 1 Throughout this document, we use ‘Research Data’ as defined by the EPSRC research data definition, which is stated as “Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.” Note the use of this definition does not restrict the subject domain, data for the humanities, arts, social science, science are all in scope, and can include recordings, images, diagrams, survey data, experimental data (including “negative data”), models, simulations etc.
- 2 royalsociety.org/policy/projects/science-public-enterprise/Report/
- 3 rcuk.ac.uk/research/openaccess/policy/
- 4 ands.org.au
- 5 ebi.ac.uk
- 6 epsrc.ac.uk/about/standards/researchdata/expectations/
- 7 arma.ac.uk
- 8 rluk.ac.uk
- 9 rugit.ac.uk
- 10 sconul.ac.uk
- 11 ucisa.ac.uk
- 12 See **Research at risk** above
- 13 repository.jisc.ac.uk/5936

Vision

John MacColl of RLUK closed the workshop in November with a summary of the delegates' views and a clear, achievable vision:

"What we want to see, looking five years ahead, is a new position accorded to data within the scholarly communications environment. We want to see data being routinely managed with the necessary articulated infrastructure (in part provided by Jisc) in place.

This will be a trusted landscape, formed of data archives at the levels of institution, region and nation, and international disciplinary archives. The skills of librarians, IT specialists and others will be required to address the challenges around capture, discoverability, preservation, storage, software and retention.

The current compliance drivers should recede into the background, expressed through a light-touch, low-cost audit approach. This should allow for the reassertion of scholarly values and a reduction in box-ticking and bureaucracy.

The data equivalent of subject/liaison librarians will be blended professionals working in all institutions, within proper career structures: these 'will be data specialists. We will provide established models of support for researchers within a rationalised scholarly infrastructure, with openness as a key principle. Every researcher will have an ORCID-style identifier; all deposit and publication routes will be easy and clear.

Data will take its place alongside publications within a rebalanced scholarly publishing economy. One of the key benefits of our sector rising to the challenge of managing research data, where commercial gain for publishers is not likely to be realisable, will be to drive positive change in the overall scholarly communications environment.

What we learn to do for data we should apply to publications. In other words, successful collaborative research data management could reform the scholarly publications market."

This document focuses on the five key areas that require action at national and institutional levels to enable us to realise this vision.

Five key areas for action - Policy development and implementation

Current state

In the last few years institutional development of research data management (RDM) policies has accelerated, spurred on in part by funder policies that mandate RDM and access. In this section, we look at institutional and funder policies and touch upon policies in relation to other stakeholders.

Often, institutions find it very complicated to develop an RDM policy and steer its adoption at the highest internal levels. Implementation can be trickier still.

Nonetheless, 40 or more universities have developed and adopted RDM policies although none would say their policies and infrastructure need no further development. Far more have yet to develop a policy, even though the drivers to develop and adopt one are becoming more numerous and more urgent. Delegates in Cambridge identified two of the most pressing:

- » Funders' requirements are likely to increase and more of them may follow the Engineering and Physical Sciences Research Council (EPSRC) example by setting out both an explicit RDM policy and a timetable for compliance
- » Publishers are more frequently requiring authors to provide access to the datasets that underpin their published research, increasing the likelihood that researchers will look to their institutions for advice and support

Interpreting what funders' policies require with regard to RDM is not always an easy matter. Partly, that is because this is such a new area, though waters are muddied further by the fact that requirements can vary from funder to funder and also between disciplines. And where projects receive funding from more than one source, their different priorities sometimes add another level of complexity.

Furthermore, it simply takes time to put infrastructure - policy, people and technology - in place.

There is a view that funders' policies are subject to regular change and that the need to focus carefully on compliance can have an adverse impact on the research process itself but, at the same time, our interviewees acknowledged that mandates are successful in galvanising institutions to prioritise change.

One prevalent view is that current policies don't build or support a reward culture; people point out that doing so would be an effective way to encourage researchers to engage willingly with the RDM process.

Another issue for funders to consider is whether a more nuanced approach would be useful, allowing for development of different requirements for different types of data; there is an appetite, too, for additional dialogue with publishers about the issues raised by open data.

Key points

Policy development

- » More work with funders is needed, both to help universities understand funder requirements fully and to influence future policy development and implementation
- » Finding ways to create a reward culture would be effective in encouraging researchers to engage willingly with RDM processes
- » Relatively few UK institutions have adopted an RDM policy yet, but that doesn't mean that no RDM implementation is taking place: some are building a policy on initial experience of implementation
- » Approaches to policy development vary and while some universities are content to mimic or adapt the policies of similar institutions, others have opted to tailor policies that reflect their own particular requirements

- » Policy drivers vary so the emphasis of policy varies too: some are aspirational and focus on best research practice while others focus more simply on complying with funders' requirements
- » A 'one size fits all' approach to policy development and adoption is not likely to be the best course of action in future
- » Given the current dynamic nature of the RDM landscape it makes sense to frame policy with an eye on possible future developments
- » The drivers that influence institutions to develop RDM policies are likely to become both more numerous and more insistent

Advocacy and awareness raising

Advocates for RDM, whether from the research office or the library, have a key role to play in explaining to researchers why they must take notice of institutional RDM policies:

- » Compliance with funder requirements
- » Benefits for themselves and their research
- » Benefits for the university

Advocacy starts when the RDM policy is adopted but it continues for the long haul; in particular, effective awareness-raising strategies will be needed to support introduction of RDM-related services.

Advisory services for researchers

Advice and guidance services for researchers (such as those that focus on helping them to prepare a data management plan for their research grant application) represent a prime opportunity to reinforce the key points of the RDM policy and explain how the university can help with looking after researchers' datasets. It is an approach that starts researchers off on the right track and it also gives RDM staff early warning of researchers' future data management needs so that they can plan active and long-term storage effectively.

While researchers generally respond well to data management plan (DMP) advisory services, institutions take different approaches to the issue. Some mandate use of their service – including the University of Manchester¹⁴, which issues a reference number to researchers who write a DMP and requires them to quote the number in order to proceed with their bid. Others make use of similar services optional, though those that do not mandate use of the service report that take-up is patchy.

Metadata catalogue

Some consensus on high level discovery metadata has been achieved and there is a need now to implement metadata to support the administration of research and research data. There has been progress and we are beginning to see agreement and uptake of some identifiers (eg ORCID¹⁵ for researchers) and vocabularies; but there is more work to be done on the legal and temporal aspects, organisational details and funder details. Further focused work on agreeing a schema and supporting implementation is a priority: it will ensure that RDM decisions can be made and policy ambitions realised.

A system that captures, manages and exposes the metadata that describes datasets is an essential part of an RDM system. Our sample of stakeholders described various approaches to metadata capture and management:

- » Some institutions look to their current institutional repository or a new one operating in parallel. Most seem to prefer a new version of their current repository software specifically for RDM purposes and they have often spent considerable effort adapting repository workflows and configuration so that they are suitable for archiving datasets
- » Others plan to use their Current Research Information System (CRIS) for the purpose although there have been reports that commercial providers of software are not always willing to develop the RDM-specific functionality that is required

- » As a third option, institutions are embarking on projects to develop new systems that meet their RDM needs but this is a practical solution only for those with in-house technical development expertise and capacity

This is not to suggest that all institutions start with the metadata catalogue. Some solve their storage issues first and most of these have yet to build, buy or develop an appropriate metadata management system. Institutions have commenced work towards shared metadata – for example, a Jisc prototype national level research data registry¹⁶ exists, drawing on data from nine HEI based catalogues and some disciplinary repositories – and some have developed their own registries of data using solutions such as CKAN¹⁷.

Archival storage

Storage is a significant issue to which institutions have devised a variety of approaches. In many ways, they are hostages to their own pasts: the ones with a centralised data centre or high performance computing centre can often move forwards quite quickly with implementation. But many have little or no suitable archival storage infrastructure and face tough decisions about whether to outsource storage or buy in solutions and manage it locally. A smaller proportion have storage capacity that may be enough for now but they face technical and organisational challenges when it comes to providing an integrated storage solution because their existing infrastructure is distributed across different faculties or multiple sites.

Long-term storage

An inadequate understanding of long-term storage requirements is a significant issue.

While some datasets will be used regularly and should be available very quickly on request, others will be used infrequently and can reasonably be stored on cheaper – probably optical – media. System back-ups may be stored in a similar way, but practical decisions about which

storage media and access configurations to choose can only be made when an institution has a reasonable idea about the demands for storage that they will face.

The issue of preservation compounds that difficulty. There is no unanimity about what kinds of datasets to keep or even about who decides. Should it be information professionals in the library or RDM team or are researchers the only people who can really know the current and future value of the data they produce? Equally importantly, how long should it be kept for? The EPSRC stipulates that datasets should be kept for ten years after the latest access; other funders may decide on different parameters.¹⁸

[1]

¹⁴ library.manchester.ac.uk/ourservices/research-services/rdm/

¹⁵ orcid.org/ and orcidpilot.jiscinvolve.org/wp/

¹⁶ dcc.ac.uk/projects/research-data-registry-pilot

¹⁷ ckan.org/

¹⁸ repository.jisc.ac.uk/5929

Five year vision

Sector-wide

Five years from now the 'mandate messiness' being experienced today should have been resolved and funders' requirements should be both unequivocal and well aligned; universities should have a very clear understanding of what funders require of them and what they need to do to achieve it.

Guidance on which data should be kept (and how long for) will be unambiguous and meeting those requirements will be both achievable and affordable, thanks to extended constructive dialogue between the university sector and research funders. Similarly, these groups will have had detailed discussions with publishers to make sure that understanding and action converge to benefit scholarly communications in general.

In five years' time the prevailing culture should be characterised by incentives and rewards rather than primarily by mandates and compliance.

University-level

Five years from now, every UK university should have an RDM policy that has been approved at the highest level, alongside a clear policy on access to data for other users.

There should be mechanisms that enable regular review of the policies and the implementation teams should include representatives from across all relevant departments who are working well together to make sure the policy is implemented correctly. High level support and a commitment to draw in specialist skills from across the institution will make that implementation a success.

First steps

Sector-wide

- » Work closely with research funders to identify and map their RDM-related requirements, analyse policies and communicate the information clearly to the university sector

- » Create RDM policy guidance with templates that universities can adapt and adopt easily

Institutional

- » Garner support at senior levels to remove obstacles and speed up progress
- » Ensure that the working groups responsible for implementation have representatives from key departments and from the research community, and make sure that a senior academic (typically the pro vice chancellor for research) leads a top level working group
- » Understand the data storage needs of researchers
- » Identify the system for metadata collection, management and storage and develop it alongside other parts of the wider RDM system
- » Ensure that the policy addresses preservation needs

Next steps

- » Achieve unequivocal positions on preservation across the disciplinary spectrum
- » Reach a consensus on what universities must do to enable reuse of the data sets they store, notably whether they need to facilitate it actively, perhaps through the provision of tools
- » Commence ongoing talks with publishers so they can participate in the RDM solution
- » Find practical ways to shift from compliance to professional rewards for researchers possibly through the Research Excellence Framework (REF)
- » Commit to longer-term provision of quality advice to institutions

Five key areas for action - Skills and capabilities

Current state

We have already touched on the fact that RDM depends on people with a whole range of complementary skills from across the organisation, but those skills are only part of the story. RDM staff also need a detailed understanding of the particular institution, the key people who work within it and the political, cultural and procedural frameworks that exist. Being able to influence the right people at the right time is reported to be of central importance.

The data equivalent of subject or liaison librarians - 'blended individuals' who will be data specialists - will need appropriate career structures of their own. Additionally, we will need to identify the key processes in RDM and the research lifecycle so that it becomes clear where existing and new specialist skills should be deployed. Our focus should be on supporting researchers to manage data in a digital environment since these are the people who know both the nature and the value of the data they produce.

Researchers too need support in developing the skills required for RDM - often this support comes from institutional RDM staff, or from more experienced researchers. Research conducted by the Knowledge Exchange¹⁹ has highlighted the key role of 'research culture' in a number of sub-disciplines as a way of encouraging researchers to develop these skills.

Of course, researchers have been involved in RDM for a long time but the current emphasis on making private data public in a form that supports reuse by others is quite new so it is here that they will need support the most. We would argue that researchers and support staff need particular help to acquire a technical appreciation of the systems that guide data management planning and data curation alongside a range of softer skills such as relationship management and management of collaborative activity and advocacy.

Most of the institutions we talked to have at least one dedicated member of staff supporting researchers and others on RDM matters, often on a fixed contract either because of the nature of the available funding or as a reflection of the uncertain nature of future RDM services in institutions.

What does an RDM all-rounder look like?

Ideally, they would have skills in:

- » Policy development
- » Business analysis
- » Advocacy
- » Project management
- » Metadata cataloguing
- » Data archiving and preservation

They would also have a good working knowledge of:

- » Data Management Planning advice and policies
- » The institution's procedures, processes and personnel (and the soft skills to get things done)
- » Relevant legal and ethical issues
- » Researcher workflows and practice
- » The IT environment

It would be a tall order to expect just one person to do all these things competently at the detailed level, but some in-depth knowledge and experience in at least a couple of these areas, plus an appreciation of the issues involved in the rest, appear to be necessary.

Almost inevitably, the person or people charged with responsibility for RDM will work as part of a multi-disciplinary team but they'll have a breadth of knowledge that enables them to coordinate progress effectively. It seems essential that they will be able to bridge divides between the library, academic departments and the IT/IS technical team.

Training

Almost all staff involved with RDM will need training. Some is being delivered already, often 'on the job', where an RDM lead will get up to speed on less familiar aspects of their role – but this is not necessarily the most efficient way to do it, especially in some skills areas. Embedding the basic skill sets can be challenging outside the core RDM team and structured training is probably the best and quickest way to develop the necessary skills. The Jisc-supported Digital Curation Centre (DCC) has already provided practical help of this kind to numerous institutions, and open educational resources (OERs) such as those developed within the Jisc research data programme have also been utilised. A third approach has been to commission other institutions to deliver training sessions.

There is a widely held view that library schools could do more to equip their graduates with RDM-specific skills, and also that Jisc could play a part in identifying the training that is needed and then coordinating the means to provide it.

Not everyone learns best in a formal (and perhaps solitary) training setting. For those who do not, shadowing a more experienced RDM person in a similar institution could be a more effective way to learn: a register of shadowing opportunities could be very useful.

Recruitment

Certainly, it is increasingly possible to recruit an experienced RDM person but it is more usual to co-opt existing personnel, especially library staff who often seem to be the most obvious choice as RDM lead.

Alongside the one or more dedicated RDM staff, universities frequently decide that a larger number of people need at least some training and basic skills that can be used to support the RDM service. That may mean they advise on DMPs, explain the RDM policy and help researchers to manage their data in a way that complies with institutional or funder policies.

Key points

- » The successful development and implementation of RDM policies depends on a wide range of skills
- » No one person is likely to have all those skills but the RDM manager will need a good understanding of all the issues and a strong set of soft skills to help in ensuring that complex projects are delivered efficiently. The ability to promote RDM and sustain networks of professionals is essential
- » Learning on the job is common and the provision of training courses by external providers is seen as both useful and effective
- » Opportunities to shadow their more experienced counterparts could be a useful alternative to formal training, or could complement it

Five year vision

Our vision calls for blended professionals – data specialists²⁰ – working within proper career structures.

This will not happen uniformly without central efforts to harmonise training - so training needs will have been identified and new courses will teach core skills and capabilities; undoubtedly individuals will then adapt these skills to the particular circumstances of their own organisation. A shared understanding of this job role in the RDM context will enable the development of relevant skills and will ensure that institutions that lack suitably skilled and experienced employees can recruit the people they need.

First steps

- » Community representative bodies should agree a list of core skills and capabilities based on the needs expressed in this report and look to see what training programmes and perhaps qualifications can be developed. Advances have already been made on the provision of online training exemplified by the RDMRose²¹ project and H2020²² also seeks to address this issue in calls related to skills
- » Jisc could play a role in identifying training requirements and coordinating the means to provide that training
- » Shadowing an experienced RDM person in another institution may be a better way to learn than formal training in isolation, so a central register of suitable opportunities should be piloted

Next steps

- » Develop training programmes, building on the materials and programmes that are already supporting skills development
- » Guide and support development of proper career structures with qualifications and job titles that are widely recognised
- » Investigate post sharing for smaller and less research-intensive institutions, potentially fostered via a national service

[1]

²⁰ The issues surrounding the skills, roles and career structure of data scientists and curators were reported in a Jisc-funded project in 2008:

bit.ly/1ICCVbF

²¹ sheffield.ac.uk/is/research/projects/rdmrose

²² ec.europa.eu/programmes/horizon2020/

Five key areas for action - Infrastructure and interoperability

Current state

Delivering the appropriate level of infrastructure at a cost that is acceptable to the institution is challenging. The policy section of this report offered some recommendations around the institutional and national frameworks required to structure such work, and identified some of the key technical solutions that need to be supported. This section focuses on the practicalities of the technologies and services involved.

What infrastructure is required?

This is not easy to define. With few basic facts at their disposal, institutions are having to make plans on the basis of their best estimations. Researchers need long-term storage but also a short-term version that enables the sharing of active data between research collaborators. Dropbox and similar services are sometimes used to do this but it is not always an ideal choice from the institution's viewpoint, notably because of the legal and ethical dimensions of sharing via cloud-based services.²³

However, it is clear that a successful technical infrastructure will have these components:

- » A system for collection, managing and exposing appropriate metadata
- » A data archive
- » Long-term file storage

Metadata catalogue

Approaches to metadata collection and management vary between institutions. Some are using existing or additional institutional repositories while others have chosen to use their current research information systems (CRISs). Those fortunate to have internal development resources at their disposal are building their own systems. In due course Jisc will provide the Research Data Discovery Service - a national catalogue building on a prototype developed with the DCC²⁴ - but, for the moment, universities must provide their own systems. That is not easy: some are concerned about the inflexibility of certain commercial

CRIS systems and also about the apparent wastefulness of each university having to do its own thing. But it is undeniable that metadata is essential for interoperability, discoverability, the provision of management information and also for making data more usable.

Storage

While the fortunate few already have high capacity data storage facilities (typically set up to service high performance computing services), most have network storage that is not up to the task of storing research data.

Some have responded by buying in new storage capacity, sometimes using external project money as seed capital. The type of storage systems they select depends on the advice of their IT/IS departments, their estimates of the capacity needed and the associated set-up and ongoing costs.

Normally, a data archive will store important or regularly-used datasets where fast access is required; back-end storage accommodates rarely-used datasets and provides capacity for backups.

The amount of free storage provided to researchers varies markedly, from one institution that provides a generous 20Tb of storage for research datasets to every research project through to the more modest (and reportedly more typical) 0.5Tb and 1Tb per researcher. Additional capacity can be requested but researchers have to pay for it, usually through their research grant. Institutions report ongoing difficulty in working out how to allocate storage effectively across different disciplines; some faculties produce more data than others but that doesn't necessarily mean they need lots more storage because they often have access to external data centres and subject-specific repositories. One university reportedly allows researchers to pool their storage allocations where they have a need for greater storage capacity.

Service level

Most institutions are providing (or plan to provide) a general RDM service but it may not meet the specific metadata or allied information needs of individual researchers, schools or faculties. This is a pragmatic solution not an optimal one - made at a time when institutions are trying to configure and launch a complex service. However, at least one university puts the onus on researchers (with advice from library staff) to provide the metadata that best suits their datasets and considers this a means of driving engagement. In general, institutions understand the benefits that attend the accurate and detailed description of datasets and they may work towards this but, for now, their priority is getting a basic service up and running.

Interoperability

Agreed minimum standards of RDM-related metadata are necessary to enable adequate discovery and to support research administration and management; although some researchers provide good metadata, others provide the bare minimum. The N8 consortium has been working on a standard set of metadata and as part of the 'Research at Risk' project are willing to hand ownership of this work to Jisc so that the standard can be more widely adopted.

Preservation-level metadata is also important and PREMIS²⁵ has an important role to play in this context. Initiatives such as CERIF²⁶ have focused on the early standardisation of key administrative metadata fields and DataCite²⁷ is another important standard that particularly addresses discovery; these all need to be taken into account. It should be noted that individual disciplines often have their own metadata standards and ontologies; it remains to be seen whether local technical solutions can accommodate these across a broad range of disciplines without making deposit workflows overly complex.

[1]

23 repository.jisc.ac.uk/592924 jisc.ac.uk/rd/projects/uk-research-data-discovery25 loc.gov/standards/premis/26 eurocris.org/Index.php?page=featuresCERIF&t=127 datacite.org/

Key points

- » The interoperation of different systems is desirable and the adoption of common metadata standards is a key method for achieving it. As well as standards such as DataCite and PREMIS there is interest in Jisc taking the lead in supporting a common metadata standard for RDM along with associated vocabularies
- » Different institutions are starting from different positions: some already have more than adequate data storage capacity while others are buying in best-estimate solutions from scratch
- » Research-intensive universities are unlikely to outsource their storage requirements but this is an attractive option for others. Overall there is strong support for shared service solutions
- » Initially, many institutions will be offering a high-level, basic service to researchers that doesn't account for disciplinary differences in metadata collection. A few will expect researchers to drive the dataset description process
- » Any RDM system will need a means to capture, manage and expose appropriate metadata. Approaches vary: some institutions are using their current institutional repository software while others use their CRIS or build bespoke systems of their own. Software such as CKAN plays a part, too
- » Often, researchers are allocated a set amount of data storage but they may also have the option to pool their allocations with colleagues. Additional storage is available on request and at a price that is usually met through the research grant
- » Archival storage is only one part of the picture – storage during the active research phase can't be neglected or collaborations will suffer

- » Institutions need more mature information management policies: this work should tie in with wider work on cyber security

Five year vision

The community wants to see data being routinely managed with the necessary articulated infrastructure, provided in part by Jisc. The current demand for a range of shared services and solutions will be met in a responsive and considered way, offering benefits for institutions and practitioners. This will be a trusted landscape formed of data archives at the level of institution, region and nation as well as international disciplinary archives. It will be a stewarded environment of scholarship, managed collaboratively by IT services, libraries, and other specialists – bringing skills that support capture, discoverability, preservation and retention, storage and software migration.

Jisc is committed to offering a range of shared services in this space, by building on existing pilots and by trialling new systems and by defining and implementing sustainable shared services. Much of this work is supported by the Research at Risk co-design project.

First steps

Today's partial, fragmented RDM infrastructure landscape is a far cry from the one we envision in five years' time. Consultation has endorsed Jisc's role in the RDM landscape, reflecting the needs of partners and providing project leadership where it is needed. So it may be useful to note, in the context of 'first steps' that Jisc, primarily through 'Research at Risk', has already begun to explore how to navigate a route between now and the vision that the community shares.

This work includes mapping an IT architecture for successful RDM implementation to help institutions meet the data curation needs of their researchers. It will highlight gaps in provision and explain products and services that are available.

Proposed work includes:

- » Building a Research Discovery Data Service
 - » Negotiating a form of national ORCID²⁸ agreement for universities to address support for researcher identifiers
 - » Working on other identifiers such as those of funders and institutions: a CASRAI²⁹ pilot has started. The current Jisc CASRAI project is defining an approach on organisational and funder identifiers and vocabularies
 - » A more coordinated approach to DOI and data citation is being considered, informed by British Library work on DataCite³⁰ DOI provision
 - » As referred to above and in previous sections of this report there is a need for more defined and agreed metadata standards for RDM. N8³¹'s initial work on this may result in a common metadata schema that can be used nationally, a situation that would be analogous with Jisc's work on the RIOXX³² Application Profile specifying core metadata for institutional repositories and funder compliance with respect to textual research outputs
 - » Clarifying where general metadata standards and disciplinary metadata intersect. Mapping the various schemas should help universities to understand the scale of the task they face in moving beyond general metadata to supporting researchers in different disciplines to use metadata that best supports their particular data
- Next steps**
- » As, above there is clear demand for national shared services for both active and long-term storage. Specifying, costing and delivering such a service is a significant task but the potential economies of scale that might result are attractive to the community

- » There is also demand for a national approach to data preservation
- » Understanding what needs to be kept (whether for compliance or for the good of scholarship) and enabling it to be reused will require new tools. Recruiting 'preservation champions' from the research community would help, as would preservation workshops to promote good practice
- » The widespread interest in a regional approach to shared services could be pursued by existing regional consortia with input from relevant national organisations when needed
- » It may be desirable to investigate development of national disciplinary archives where they don't already exist, and Jisc and RCUK could usefully consider this
- » Services to provide usage statistics could be developed: the information would support advocacy and also be valuable for management information purposes
- » A national data management planning registry may help the HEI community to plan capacity and analyse their progress

[1]

- 28 orcidpilot.jiscinvolve.org/wp/2015/02/03/next-steps-for-orcid-adoption-orcid-consortium-membership-for-the-uk/
- 29 jiscasraipilot.jiscinvolve.org/wp/
- 30 bl.uk/aboutus/stratpolprog/digi/datasets/datacitefaq/faqhome.html
- 31 The N8 research partnership of northern England universities: n8research.org.uk/
- 32 scholarlycommunications.jiscinvolve.org/wp/2015/01/26/launch-of-riox-xx-application-profile/

Five key areas for action - Incentives for researchers and support stakeholders

Current state

Even though aspects of RDM are now 'necessary' this does not mean that they are demand-driven.

Demand from researchers for data archiving solutions is limited, but the funder-related drivers are increasingly a fact of life. RDM professionals will need to overcome researchers' reluctance to offer up research data or, better still, enthruse them so that they engage willingly with the RDM process. Our five year vision for RDM aspires towards a future in which the current compliance drivers recede into the background (or at least, have a lighter touch), allowing for an academic-led reassertion of scholarly values.

Key points

Incentives for researchers

- » Compliance alone will not result in researchers embracing RDM willingly
- » Unfortunately, the benefits of RDM and long-term storage are hard to sell to researchers and the few incentives that exist for them are insignificant compared with 'sticks' such as funder and publisher requirements or institutional mandates. It isn't surprising that some researchers respond to an authoritarian approach by doing the barest minimum with respect to provision of metadata
- » Funders mandate the archiving of datasets however researchers fear that costing it into their research proposals may make them uncompetitive. Funders could do more to reassure researchers that bids with identifiable RDM-related costs will not be disadvantaged
- » It would be useful to achieve greater clarity about what RDM-related costs can be recovered via funders' grants
- » As we have already noted, researchers need explicit, meaningful rewards for engaging effectively with RDM, either through the REF or in the form of career progression within their institution. Current reward

structures are sometimes seen as too focused on high impact journal publications

- » And again, a greater focus on the value that effective RDM brings to the publication process would also be useful. Already some publishers and journals require authors to provide access to relevant underlying datasets and this should be presented in a positive light as adding value to the article and also potentially driving up citations. Researchers could gain measurable kudos from publishing their data if there were more data-focused journals and relevant metrics
- » General advocacy would highlight the benefits to the wider research community and society of opening access to data. The prospective benefits of open data and data mining may be important motivating factors
- » RDM services that provide advice to researchers who are completing a data management plan have a good opportunity to advocate the value of effective RDM throughout the data life cycle
- » Making data more shareable would encourage a cultural change in which reusing other people's data becomes more common in more disciplines, reducing duplication of effort
- » Finding ways to offer download information and other statistics will encourage researchers to engage with RDM
- » A 'data fellow' could coach researchers in publishing data and building collaborative networks, and there should be career-related rewards for those who do carry out such coaching
- » **But of course, the strongest incentive should be that RDM makes it easier for researchers to do their work**

Incentives for support stakeholders

Support stakeholders are those within an institution who play a role in RDM services. Usually that is part of their paid job, but other incentives apply:

- » RDM is professionally challenging and offers people a chance to 'make a difference'
- » Running a pilot project will bring institutional resources forward to provide sustainable RDM services
- » As librarianship changes the library profession is highly motivated to get involved in new areas; they have a keen professional interest in providing a service and also in personal development
- » The importance of the RDM role is increasingly understood and its growing visibility within the institution gives those working in the role a greater sense of engagement
- » A key incentive is seen to be enabling their institution to comply with external requirements
- » Some say that getting their services used is an important incentive, as is the satisfaction of foreseeing and forestalling problems such as data protection issues and information security
- » Local awards could be given to RDM managers who are doing well and there could even be league tables

Five year vision

Five years from now easy access to data will be the norm and not doing RDM well will be tantamount to research malpractice.

The scholarship value statement will need to be broader than it is today.

First steps

- » Effective promotion of the standard method for citing datasets and encouraging researchers to reuse and cite other people's datasets will both be important. We need to help people move on from traditional ways of working
- » Organisations such as HEFCE and RCUK could play a pivotal role in recasting institutional values and encouraging inclusion of an explicit RDM focus in funding and research reviews. Engagement with these organisations is therefore an immediate priority
- » Searching out researchers who are open to new ideas and engaging them as role models and opinion leaders will be a powerful approach
- » Researchers are understandably more open to the idea of their university looking after their data when they have suffered data losses in the past, so messages focusing on avoidance of data loss might incentivise some. We should identify and capture their stories
- » RDM systems must be tightly integrated with other institutional systems to make data deposit as easy as possible for researchers; aspects of the 'Research Data Spring³³' work address this issue
- » An analysis of the savings made possible by reusing data rather than creating new work would generate powerful messages about the value of making data available for reuse
- » Offering free storage incentivises researchers to engage with the RDM process

Next steps

Creating the necessary changes in the prevailing culture presents a significant challenge. We need to make sure that these three things all become straightforward enough to constitute standard practice:

- » Data management planning
- » Effective data management throughout the active research phase
- » Data curation that facilitates data reuse by others

Doing so will be for the good of scholarship as a whole, and it will require constructive liaison between sector organisations and research funders.

[1]

33 jisc.ac.uk/rd/projects/research-data-spring

Five key areas for action - Business case and sustainability

Current state

The lack of certainty about what it will cost an institution to set up a robust RDM system and then sustain it into the future is a significant problem. Knowing how much storage to budget for is a particularly intractable difficulty: buying expensive machines and the capability to support them is a risk when no-one knows how readily or rapidly researchers will deposit datasets or even what size those datasets might be.

The cost of employing staff to facilitate the RDM development process is another major issue. In some cases fixed term external appointments have been made with project money; in others existing staff, often from the library, have been redeployed while in other cases staff restructuring has allowed for the appointment of a new RDM person.

The development of business cases has sometimes kick-started the RDM process, by unlocking enough money for a pilot study or service and the short-term employment of personnel. However, we heard that not many business cases have succeeded in releasing the full investment requested. In such cases a second business case might need to be made for a full-scale service. It is difficult to generalise: while we learned of instances where funds were difficult to access, there were others where the whole process was straightforward and painless. Top level support for RDM within institutions is clearly vital to the approval of business cases and so, too, are policies and skill in selling the business case to the various key internal stakeholders.

It is important for the RDM business case and the university strategy to be in alignment. RDM should be a normalised part of the overall scholarly communications environment, something that universities internalise as part of their core business processes. At our workshop in Cambridge we asked a set of questions designed to highlight the issues that RDM managers need to think about when formulating a new business case or reviewing an existing one. We wanted to know:

- » What evidence do you have of the need for RDM; has the university failed to win research funding on the basis of an inadequate DMP or RDM infrastructure?
- » What is the risk of not doing RDM in terms of, for example, data loss or an increasingly uncompetitive position against similar universities?
- » How scalable is the proposed service, noting that volumes of stored data are likely to increase over time?
- » What is the need for staffing and should they be focused on RDM full-time or have it as part of a portfolio of responsibilities?
- » How much will software, storage and associated services cost and how might this change over time?
- » To what extent is it possible to cater for different researcher needs based on disciplinary norms?
- » What is the cost of advocacy likely to be?
- » What is the preservation strategy for data and software, what is it likely to cost and how frequently will the preservation strategy be reviewed?

Costing models

Many institutions are unclear about the best approach to take on costing so are keen to know how peer institutions have approached the issue. They need advice on the validity of their business cases. Some have shared their approach with the DCC but usually only on the understanding that the information is not made generally available. Others have participated in the 4Cs³⁴ (Collaboration to Clarify the Costs of Curation) project.

The finding that institutions appear unwilling to share detailed costing information (perhaps because of the absence of this information) adds urgency to the requests that Jisc frequently receives to conduct costing case studies and

develop pro forma costing models: part of the 'Research at Risk' project will respond to these. These models may be adopted or adapted by institutions or may simply be used as a benchmarking or validation tool. Institutions want reassurance that they are investing broadly the right amount of financial and other resources compared with similar institutions. Of course, each institution will start from a different baseline in terms of existing skills, capabilities and storage infrastructure, but they still want to know that their approach to funding the service (and recovering costs where possible) is a valid one.

Approaches to funding

It is clear, then, that there is no common approach to funding RDM. Neither is it clear that the approach taken at the outset by an institution will necessarily persist into the future. The most straightforward cases are those where the institution pays the costs of RDM from either its central budget or the research budget. In others, the budgets of faculties are top-sliced to provide funding for RDM. Sometimes the costs of funding the staffing element of the service, the metadata catalogue and related development are found from the core library budget and there are also cases where the university funds the RDM advisory service centrally with the costs of storage coming from individual faculties. In other situations capital expenditure is used to buy storage machines with the intention of charging other related costs of running the service to the funding councils.

The cost recovery route for data storage from funders may be difficult (and costly) to administer and it is not yet clear how the charge-back mechanism will work; we need greater clarity. Institutions are aware that EPSRC will not allow 'double dipping' and are being careful to avoid doing so. Another approach is for the institution to bear the cost of a certain amount of storage per researcher or research group; if they need more, this must be specified in a grant application and the relevant costs recovered from the funder. Again, this will involve the use of an administrative process if the recovered funds are to find their way reliably into the RDM budget pot.

Sustainability

In two to three years' time there may be more (and possibly cheaper) storage options, perhaps through the mechanism of shared services. But for now, it is difficult for institutions to plan too far ahead with any certainty. In some cases investment for big projects for a particular university has to be planned on a full cost basis for a ten year period at net present values, which presents a major problem.

Across the board preservation appears to be, at present, the poor relation of the RDM family. It seems that most institutions do not yet have a full grasp of preservation issues in relation to datasets: the skills required are in short supply and the issues are complex and potentially costly. No-one can say which research datasets will be required after ten years, which will be popular, which may be deleted - this all has a bearing on costs. Some institutions have not fully considered how to sustain RDM for the long term and some think that the process of recovering costs from funders may be too difficult.

Regarding the potential benefits of RDM, compliance with funders' requirements may sustain current levels of revenue or even increase them if competitor institutions are less able to demonstrate compliance. And soon, research-related savings derived from reusing rather than newly creating data may become quantifiable. Download or reuse metrics may demonstrate increased impact that may in time translate to a better competitive position for the institution. And a growing acceptance of the value of data sharing may lead to new collaborations with new partners that unlock new sources of funding for an institution. Where financial benefits such as these can be quantified they can help to justify an institution's investment in RDM. These variables could all scale to regional and national levels allowing the benefit of RDM to the UK to be demonstrated to those that fund the system.

currently struggling, particularly when it comes to planning and costing an appropriate RDM system.

In five years' time we would hope that all UK universities will have RDM systems and services in place, operating successfully and increasingly sustainably for the good of the institution but also for scholarship more generally. Universities will be compliant with clearer, harmonised funder requirements, they will have a clearer view on the financial and marketing benefits of their RDM investment and researchers will be benefiting from more tangible rewards for sharing data.

First steps

- » A generic RDM policy template would help individual universities build a formal business case for RDM investment in to their developing policy. A selection of these can be developed centrally, reflecting the different nature of universities including their reliance or otherwise on funding from the research councils
- » Similarly, a cost template would help university managers to understand what similar institutions are spending on RDM and how they are finding and allocating the money
- » Creating a common list of the benefits that institutions and their research communities can realise through investment in RDM would be helpful. Work - drawing on the Keeping Research Data Safe³⁵ project - to develop ways to measure the benefits would support the argument
- » Work is needed with research funders to clarify the costs that can legitimately be recovered from them
- » Sector organisations such as the six sponsors of the Cambridge workshop are well placed to make the case to universities for RDM to be regarded as a routine and very valuable part of the scholarly communications process

Key points

- » Approaches to funding RDM services and infrastructure vary hugely
- » In general there is uncertainty about the storage capacity that is required now and in the future
- » There is strong desire for standard costing models supported by case studies that institutions can adopt or adapt to their own circumstances
- » Uncertainty remains about how much of the cost of the RDM service and infrastructure will be recoverable from funders, together with some apprehension about how difficult the process will be
- » The sustainability of all aspects of RDM is something that has still to be considered in most cases: there is so much uncertainty about the issues that they have been put to the back of the queue, at least for now
- » Good quality management information should help senior university managers to justify investment in RDM infrastructure and services

Five year vision

Our vision statement looks beyond the issues of how individual institutions choose to finance RDM to the wider system and it speaks to a high level ambition for the sector. As John MacColl said: "One of the key benefits of our sector rising to the challenge of managing research data, where commercial gain for publishers is not likely to be realisable, will be to drive positive change in the overall scholarly communications environment. What we learn to do for data we should apply to publications. In other words, successful collaborative research data management could reform the scholarly publications market."

Rising to the challenge of RDM will make the sector's wider aspirations for scholarly communications more achievable but this is an area where many universities are

- » Business cases may well suffer setbacks when the need for new staff is identified, especially if the posts are permanent. A good short-term tactic may be to reorganise the responsibilities of existing staff so that they can be trained to take on some of the RDM load
- » National shared services and national approaches for components like storage, preservation, metadata standards and unique identifiers will take time to develop but if organisations with a coordinating role can offer early reassurance that this work is in progress, it will give universities greater confidence that they won't have to tackle all these issues themselves

Next steps

Not surprisingly, much of the activity so far has focused on research-intensive universities since these have most at stake in terms of compliance with funders' policies. Revenue from the research councils represents a much smaller proportion of total revenue for a sizeable group of other universities and there is less information available on how they will respond to the RDM challenge. Advocacy from national organisations about the benefits to scholarship may be helpful in persuading these institutions to engage actively with RDM.

We need more clarity on several issues:

- » The costs associated with keeping RDM systems up to date as new technologies emerge
- » The costs associated with preserving data longer term and the challenges of providing tools to enable reuse of those data
- » Practical ways to collect local and national level statistics about data deposit and reuse, to support advocacy, service planning and management information

[1]

35 beagrie.com/krds.php

Moving forwards

This report identifies challenges surrounding RDM for the UK's academic institutions and offers some practical steps for the future – but it is just one component of wider activity.

The UK's HEI community knows itself to be both resourceful and innovative, and to plumb this well of innovative thought Jisc has recently launched Research Data Spring³⁶, a project that encourages individuals and groups with an interest in research data to come up with new ideas and solutions to common problems.

ARMA, Jisc, RLUK, RUGIT, SCONUL and UCISA have been instrumental in seeking the views of the community and striving to understand in detail the road that institutions need to travel as they plan and implement RDM systems and services that are fit for purpose. We've identified where the community wants to be in five years' time with respect to RDM, based on a number of consultation processes and the views that have been expressed to us are represented in this report. They now need to be considered for action.

The 'first steps' sections of the report set out where work is already in hand, under active consideration or where useful outcomes can be achieved sooner rather than later. The 'next steps' will require more thought, planning and some serious conversations about how any actions should be pursued. If we are aiming ultimately to develop a full roadmap, the Cambridge workshop helped to formulate a commonly agreed destination and this report provides the waypoints – ideas from which a series of activities or work packages may be designed to facilitate the journey. That journey itself is likely to be as challenging as it is professionally rewarding; if it is hard going at times, it will help to remind ourselves that the ultimate prize will be enhanced scholarship in the UK's universities.

[1]

³⁶ Within the Jisc Research Data Spring ideas around research data deposit and sharing tools; data creation and reuse by discipline; research data systems integration and interoperability; research data analytics and shared services for research will be developed at a sandpit workshop and successful teams will receive funding. In particular this seeks to address some of the integration issues that exist and are also raised in the roadmap. This is just one element of the wider 'Research at Risk' project and the cohesive action towards shared solutions that the roadmap highlights will be the majority of the Jisc 'Research at Risk' initiative's focus.

Share our vision to make
the UK the most digitally
advanced education and
research nation in the world
[jisc.ac.uk](https://www.jisc.ac.uk)

Jisc

One Castlepark
Tower Hill
Bristol, BS2 0JA
0203 697 5800
info@jisc.ac.uk