

How to Develop Research Data Management Services - a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte

Please cite as: 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. Available online: <http://www.dcc.ac.uk/resources/how-guides>



Digital Curation Centre, March 2013
This work is licensed under Creative Commons
Attribution BY 2.5 Scotland

How to Develop Research Data Management Services - a guide for HEIs

1. Introduction

The purpose of this guide is to help institutions understand the key aims and issues associated with planning and implementing research data management (RDM) services. It explains the components and processes of RDM services and describes the roles and responsibilities of those who will deliver and use them.

Many higher education institutions (HEIs) have recognised the need to develop RDM services and are currently engaged in this activity. Over 40 UK universities have been involved in developing RDM services within the Jisc MRD programmes¹ and DCC Institutional Engagements². This guide summarises their lessons and emerging approaches, and will be of help to other institutions beginning to address these issues.

The guide has been written for anyone working in an HEI who has an active stake in the generation, management or sharing of research data. This includes university management, support and administrative services, and researchers. Indeed, there are many other functions and roles supporting the research data agenda to which we hope this guide will provide assistance.

2. Why develop research data management services?

There are many benefits and drivers that prompt the development of RDM services. Effective management of data promises rewards throughout and beyond the life of a research project. Whilst it is important to ensure that data are discoverable, accessible and intelligible to enable long-term reuse, such values are equally crucial during the research-active phase. For the researcher, the perception of data as an instrument of research and new knowledge can be transformational. Well-managed data lead to higher-quality research, increased visibility and the consequent benefits of enhanced citation rates.

The increasingly collaborative nature of research is a pressing argument for RDM services. Researchers need to exchange data across diverse platforms and demand effective systems to store, access and share data securely across multi-institutional

research teams. The expectations of research funders have also driven work in this area. The EPSRC Policy Framework on Research Data³ in particular has instigated the development of RDM services within many UK HEIs. Universities play a key stewardship role in curating the outputs of research. Research data are an asset, bringing benefits and impact for the institution as much as for the researcher. In order to reap these benefits, effective systems and support services need to be in place.

3. Roles and responsibilities

The range of skills and knowledge needed to deliver RDM services is dictated in a large part by the individual phases of the research project lifecycle. Consideration of these phases provides some indication of the activities that will be required by a full support service:

- **at pre-award:** assistance with the preparation of data management plans, including guidance on costing data management activities and the expert use of online tools;
- **throughout the project:** advice on data documentation, formats and standards to enable reuse; guidance on storing, managing and analysing data to achieve regulatory compliance and best practice; advice and/or provision of research data storage facilities that meet the needs of a wide range of data types, platforms and access needs;
- **post-project:** advice on selecting data of long-term value; support to make research data visible and/or available to defined audiences; help for researchers in deciding how to archive data at the end of a project (or at any other appropriate point).

Meeting all of these needs cannot be delegated to a single unit. In the complex and frequently distributed organisational environment of an HEI, the roles and responsibilities for defining, enabling and delivering RDM services are generally shared across three groups: university management, support and administrative services, and researchers.

¹Jisc. *Managing Research Data programme 2011-13*, URL: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

²DCC. *Institutional Engagements webpage*, URL: <http://www.dcc.ac.uk/community/institutional-engagements>

³EPSRC. (2011). *Policy Framework on Research Data*. Retrieved 12 March 2013, from <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>

3.1 Management

The principal role of university management in the introduction of RDM services will be to ensure proposed services are desirable, achievable and sustainable, and if so, to give clear, informed and unequivocal support. Equally challenging for university managers is the need to treat RDM services as a serious investment in infrastructure within the long-term institutional business planning process.

Specifically, the principal responsibilities to be met by senior management are to:

- provide a champion, at Pro Vice-Chancellor (PVC) Research level or equivalent, to act as influential advocate and to chair steering/working group business;
- establish a representative, balanced and appropriately equipped steering/working group that will reflect the interests of essential stakeholders;
- consider, comment on and eventually approve proposals, plans and strategies, including the endorsement of budgets and organisational restructuring;
- advise on the higher-level strategic issues that must be addressed during service design;
- ratify a policy that articulates the core RDM principles and acts as a framework for guidelines and service design.

3.2 Support and administrative services

The support teams that will deliver RDM services may typically be categorised as the library, information technology, records management and research administration functions, although this list is not exclusive. For all of them, managing research data is likely to be a relatively new challenge for which the responsibilities and practices have yet to be firmly established. Those traditionally engaged in information management and computing services will, however, be customarily recognised as the groups best fitted to lead in the identification of requirements, standards and solutions. Research administration, involving the management of grants and contracts, commercialisation and the support of innovation, comes in diverse shapes across the sector. In most cases it plays a key role as the link between researchers, management and funders. It is important to recognise that in many institutions these groups will not have previously worked as a cohesive unit or partnership. To deliver RDM services they will together provide the effort to:

- establish an RDM team to undertake the actions defined by the steering group;
- undertake analyses of policy requirements at a national, funder and institutional level;
- identify the nature of the institution's data assets and data management practices;
- develop and implement proposals, plans and budgets for the technological and human infrastructures necessary to deliver RDM services;
- retrain, reorganise and otherwise acquire the skills for providing effective RDM services;
- plan and undertake a programme of advocacy to promote the key aspects of effective research data management, explaining in universally accessible terms its obligations, benefits and the services anticipated;
- facilitate training opportunities for managers, support staff and researchers;
- as traditionally independent units, reorganise into a partnership to deliver a seamless RDM service.

3.3 Researchers

As the creators and users of research data, researcher engagement is crucial in the development of RDM services. Any service provision needs to be based upon a close understanding of research, its patterns and timetables, motivations and priorities. This cannot be achieved without a commitment by the research community to contribute to the definition of service requirements. Without their active involvement and support the success of an RDM service is bound to be limited. Whilst management will define expectations and support staff will deliver services, it is the responsibility of researchers to:

- ensure their views are represented by contributing to steering/working groups;
- collaborate in the gathering of requirements and the testing of solutions and methods;
- clearly articulate - in terms of their data creation, use and management - the particular requirements, opportunities and obstacles encountered within their disciplines;
- champion the adoption of approved methods and services within their communities.

4. The process of developing services

4.1 Acknowledge differences in organisational culture

In order to support data management and sharing, HEIs need to coordinate across a diverse range of actors and processes to deliver the necessary technological and human infrastructures. A rigid model for this cannot be prescribed since individual organisations and cultures occupy a spectrum of differences. However, it is feasible to describe a set of common components, process and services, which is the focus of this How-to guide.

Before selecting from any menu of service components, it is important to scope the institutional context, to explain and agree why RDM services are needed, to understand what specific challenges the services are expected to resolve and what, essentially, such services should deliver. Strategic decisions will need to be taken. For example, you need to define the role played by national and international services, which may be regarded as alternatives or as complementary to institutional services. How an HEI responds to these and similar questions will be determined by a heady mix of its prevailing research culture, its mode of governance, the scale of its competitive aspirations and, not least, by the measures at its disposal to predict and manage resources within the broader profile of its operating plan.

Convening a group of interested stakeholders is a useful first step. This should be led by a senior champion such as the PVC for Research or an esteemed researcher with cross-discipline authority. The group should be broadly based, and is likely to include representatives from the library, IT services, the research office, ethical advisors, archive services, legal experts and the research community. An RDM project team is also needed to lead the development of services.

4.2 Undertake requirements analysis

It is necessary to gather requirements and undertake a gap analysis to inform the development of RDM services. You need to identify any gaps between your current position and where you hope to be in the future in order to plan the activity needed to make

the transition to a fully functional service. Be aware of any contingencies between steps so that a logical sequence of activities can be established. It is also crucial to assign clear responsibilities and to set measurable indicators for success.

Two DCC tools support this work and have been adopted and adapted successfully by universities engaged in the Jisc MRD programme and the DCC's Institutional Engagements.

The Data Asset Framework (DAF)⁴ is a survey and interview-based methodology to investigate research groups' data holdings and how these are managed. Questionnaires and interviews generally cover the range of activities involved in the curation lifecycle to identify issues and expectations of improvement in support. DAF has been piloted in a number of contexts and has been the subject of several case studies⁵.

The Collaborative Assessment of Research Data Infrastructures and Objectives (CARDIO)⁶ tool aims to help establish consensus on RDM capabilities and identify gaps in current provision. Institutional preparedness is assessed using a capability model adapted from the 'three legged stool' approach used by Cornell University's digital preservation programme.⁷ CARDIO users rate existing provision in three areas - organisation, technology and resources - and come together to agree their ratings and to prioritise action. The tool can be used online, in person or in a combination of these.

Requirements-gathering and gap analysis complement one another. Many institutions have begun the process by consulting with academic staff in two or three research groups or departments in different schools or faculties. It helps to involve researchers whose experience spans a range of funders, career stages, research disciplines and data types. Workshops or focus groups that involve relevant service providers alongside the academic staff from these studies are particularly useful fora in which to frame the issues that emerge, and will serve as platforms for the consolidation of findings and consequent action planning.

4.3 Pilot services before implementation

Once requirements have been identified and you have a plan of action, consider existing tools and service models that may be reused. Each of the topics in section 5 points to examples from other universities. Lots of lessons have been learned by early adopters

⁴Data Asset Framework website, URL: <http://www.data-audit.eu>

⁵Jones, S., Ball, A. & Ekmekcioglu, C. (2008). *The Data Audit Framework: a First Step in the Data Management Challenge*. *International Journal of Digital Curation*, 3 (2): 112-120. doi:10.2218/ijdc.v3i2.62

⁶CARDIO website, URL: <http://www.dcc.ac.uk/projects/cardio>

⁷Kenney, A. & McGovern, N. (2005). *The Three-Legged Stool: Institutional Response to Digital Preservation*. *II Convocatoria del Coloquio de marzo*. Cuba. Retrieved 12 March 2013, from http://www.library.cornell.edu/iris/dpo/docs/Cuba-ark-nym_final.ppt

and you can benefit from their experience. Many of the outputs are licensed for reuse or adopt open source technology to enable other HEIs to adapt them to suit their context.

When developing potential solutions it is critical to engage the user community and pilot services to ensure they are fit-for-purpose. At the University of Edinburgh, for example, the steering group has established three preliminary pilot studies: one on data management planning, focusing primarily on the use of the DCC's DMP Online tool; one on research data storage exploring the extension of the ECDF-NAS service⁸, a chargeable large filestore for research groups; and one on the DataShare repository⁹. Once you know the services are fit for purpose and meet a broad range of use cases they can be adopted across the institution. A phased roll out may be appropriate to ensure uptake can be supported.

Developing services – a summary of key actions

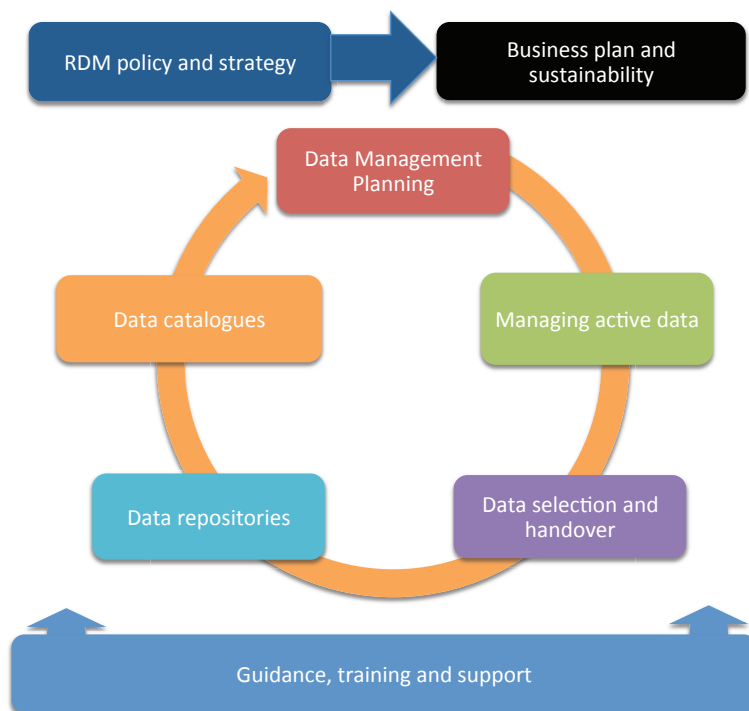
- Assemble a broad-based steering group, with academic involvement, led by a senior research representative e.g. PVC for Research;
- Appoint an RDM team with appropriate responsibilities and competences to undertake actions;
- Scope the RDM status quo, including gaps in service and support, to identify requirements;
- Design services that meet local and external drivers, and suit the organisational culture;
- Pilot services to ensure they are fit for purpose before wider roll out.

5. Components of an RDM service

In order to support effective data management and sharing, an institution needs a coherent strategy and suite of services. The following section proposes a number of components to be addressed when delivering RDM services, together with a description of the roles and responsibilities of those who may deliver and use them. Practical guidelines are provided to suggest how each component can be addressed and a series of case studies is forthcoming to provide more detailed examples of approaches that may be taken.

The diagram below attempts to visualise the different aspects that need to be addressed.

Components of research data management support services



⁸Edinburgh Compute and Data Facility (ECDF) Networked-attached Storage service website, URL: <https://www.wiki.ed.ac.uk/display/ecdfwiki/ECDF+NAS+service>

⁹Edinburgh DataShare repository, URL: <http://datashare.is.ed.ac.uk>

5.1 RDM policy and strategy

Developing a strategy

Having an overarching strategy is essential to ensure that RDM services develop coherently. It should outline key objectives and the stages of work planned over a set period in order to realise them. There are three key steps to defining your strategy:

1. understand your current position;
2. define where you want to be in the future;
3. map out a programme of activity to make this transition.

Requirements analysis exercises as outlined in section 4.2 are a critical step in defining your strategy. In order to take stock of the current situation, you need to be aware of the context in which you are working. What internal and external factors influence research data management and sharing, and which have the greatest implications for you? These could include codes of conduct for research, funder policies, national and international legislation, and collaborative agreements that necessitate the sharing of data across institutional boundaries. You should also be aware of your institution's mission statement so that you can map RDM activities and benefits to it as a means of encouraging support.

Existing examples can provide pointers to help you get started. Monash University, for example, has shared its *Research Data Management Strategy* and licensed it for reuse by others¹⁰. The DCC has also provided a series of blog posts to help institutions develop EPSRC roadmaps and several examples have now been made public¹¹. Typically these roadmaps list the expectations to be met and perform a gap analysis to identify activities needed over the coming years. It can be useful to think more broadly than EPSRC compliance alone, since most research funders have expectations in terms of RDM. The University of Edinburgh roadmap, for example, organises planned work under four key areas: data management planning; active data infrastructure; data stewardship; and data management support. A number of objectives are listed under each area of work together with concrete actions, deliverables and target dates¹².

Developing a policy

Extensive consultation is critical when developing a policy; policy development is an effective outreach activity in its own right. You need to be aware of the

roles different stakeholders play and what their issues and needs are to ensure that the policy is desirable and realistic. Eliciting feedback and involvement throughout the development process will ensure the policy is fit for purpose. Remember to keep things simple and use clear language and concepts that speak to the people who will be expected to apply and support the policy.

Existing examples can give you ideas of what to include. The DCC collates a list of UK institutional RDM policies¹³ and several useful examples are also available from overseas. Agreed guidelines, such as the UK Research Integrity Office's *Code of Practice for Research*¹⁴, also provide a useful base as they outline common expectations for the collection, use, storage and retention of research data. The DCC has produced a policy briefing, which outlines UK funder requirements, the approaches taken by different UK universities and considerations to make when developing institutional RDM policy¹⁵.

Once you have a first draft, review this with a small pilot group (e.g. researchers, management and staff from key services) to make sure it is understandable and covers the key points. The policy will then need to be ratified by the University's governing bodies, a process that will likely require several iterations and may take some considerable time. For this reason it is useful to keep the policy brief, focusing on high-level principles. Accompanying guidance is essential to aid implementation; however, this is likely to be more fluid, needing to be updated regularly as the supporting services develop, so it can be useful to maintain this separately.

The most challenging task will be implementing the policy as this is likely to imply significant modification or development of infrastructure and changes in working practice. The approach taken at the University of Edinburgh is to run pilot studies that trial implementation in a number of areas, using these as examples to roll out the emerging practice out more widely. Finding ways to incentivise adoption, for example through career progression, may prove to be especially worthwhile.

Coordinating data policy and strategy

For a number of HEIs, the creation of a research data management policy has been selected as the first step in the process of defining a strategy. Policies provide clarity on what is expected by the institution and who is responsible for which activities. Research data management involves many stakeholders so a coherent vision expressed through a policy can be a

¹⁰Beitz, A., Dharmawardena, K. & Searle, S. (2012). *Monash University Research Data Management Strategy and Strategic Plan 2012-2015*. Retrieved 12 March 2013, from <https://confluence-vre.its.monash.edu.au/download/attachments/39752006/Monash+University+Research+Data+Management+Strategy-publicrelease.pdf?version=1&modificationDate=1334289180000>

¹¹DCC. List of RDM roadmaps produced by UK universities, URL: <http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps>

¹²University of Edinburgh. (2012). *Research Data Management (RDM) Roadmap August 2012 – January 2014*. Retrieved 12 March 2013, from <http://www.ed.ac.uk/schools-departments/information-services/about/strategy-planning/rdm-roadmap>

¹³DCC. List of RDM policies produced by UK universities, URL: <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>

¹⁴UKRIO. (2009). *Code of Practice for Research: Promoting good practice and preventing misconduct*. London: UK Research Integrity Office. Retrieved 12 March 2013, from <http://www.ukrio.org/what-we-do/code-of-practice-for-research>

¹⁵Jones, S. (2011). *Research data policy briefing*. DCC. Retrieved 12 March 2013, from http://www.dcc.ac.uk/webfm_send/705

useful way to coordinate the broad range of interests to be found in an HEI and to provide a framework of overarching governance. A policy can also provide leverage to unlock resources for infrastructure development, making implementation more feasible.

The best time to release a policy is a moot point. Fears have been expressed that approving a policy prior to the development of infrastructure and support services may lead to an eventual gulf between aspiration and the realities of resourced and achievable implementation. Others have found a policy to be a useful starting point, a means of gaining traction and presenting an aspirational view that will motivate and guide service development. Be mindful of your institutional culture to determine the best approach.

This interplay between policies and strategies is worthy of note. In cases where a policy has been developed first, the subsequent strategy will need to provide a more detailed roadmap for implementation. Other institutions, such as the University of Bath, have developed a strategy first and addressed policy subsequently, treating it as a supporting component of service development.

Developing RDM strategy and policy – a summary of key actions

- Understand your current position and where you want to be to define your strategy;
- Map out a programme of activity to deliver infrastructure and services;
- Draft a policy based on external drivers and local context to establish your core RDM principles;
- Ratify the policy then undertake advocacy work and pilot studies to aid implementation;
- Consult broadly and iteratively throughout to gain consensus and secure support.

5.2 Business plans and sustainability

For many institutions the development of RDM services represents a wholly new enterprise, often requiring significant organisational and behavioural change. To ensure that these changes are sustained requires the creation – and formal acceptance – of a long-term business plan. A business plan should set out objectives, predicted costs and planned expenditures, resource deployment and enhancement, the forward change programme and anticipated

benefits. Given the uncertainty of levels of long-term public funding it is recommended that a phased approach should be taken, with the plan stepped over three, five and ten years. A set of three- to five-year rolling phases would match the kind of planning window applied to the operating plan in many institutions.

Such a transitional plan will enable you to set out your long-term plans without requiring immediate commitment of potentially considerable cost. Start small and plan for growth. Securing agreement to proceed with the initial steps, and having a clear statement of your long-term goal and its cost implications will ensure that a working service can be established that is less likely to meet with resistance in future years. If you can achieve a trade-off with existing services in those early years, so much the better. For example, if the service is presented as a reconstitution of the library service to meet the changed needs of research it will not be viewed as a wholly new expenditure and will be applauded as a cost-neutral enhancement that also brings new career opportunities to the staff involved.

It is crucial for the RDM services business plan to reflect the institutional mission, as expressed in its strategic plan, since this will give it legitimacy. Your RDM services strategy and roadmap will furnish your service objectives but in the business plan you will need to assign known or forecast costs, together with an indication of the financial year in which expenditure will occur. Remember that capital costs (one-off equipment or building costs) are normally budgeted annually and are not recurrent, whereas the maintenance of those capital items, along with staff costs (which will include overheads and incremental drift) form part of the institution's revenue budget. It may be easier to argue for capital expenditures, which can be allocated or phased strategically across several years, whereas recurrent expenditures will come under constant scrutiny according to the ebb and flow of institutional income and expenditure across the board.

It is equally important that you seek local guidance on interpretation of the Transparent Approach to Costing (TRAC¹⁶) guidelines and Full Economic Costing (fEC), so that your estimates do not prove insufficient and your business plan conforms to institutional principles.

Your business plan is your opportunity to win management approval and agreement that resources will be committed to ensure the service can be sustained. You should therefore describe what returns on investment are predicted, which for RDM services are likely to be expressed in terms of improved research impact, a more effective and cost-efficient research process, improved opportunities for new and more research, or increased funding. Undertaking a benefits analysis, perhaps using the KRDS Benefits

¹⁶J M Consulting Ltd. (n.d.) *Transparent Approach to Costing (TRAC) Guidance*. HEFCE. Retrieved 12 March 2013, from <http://www.jcpsg.ac.uk/guidance>

Analysis Toolkit¹⁷, will strengthen your case, as will reference to institutional examples that enable you to scope your analysis¹⁸.

Creating the business plan will require you to address the issue of charging for services. How much of the RDM services, for example, can be charged to individual research projects and how much will be corporately funded? How centrally managed will the service be and how far devolved? Much will depend on the tradition and culture within your own institution but some aspects of the RDM services such as storage or training will at least require strong central coordination, and the plan should be explicit as to how this will be managed. With regard to the use of project funding, at least two of the major funders have indicated that up to ten percent of a research grant can be used for the purposes of data management; however, whether this includes investment in long-term infrastructure or is restricted to in-project activity has still to be confirmed¹⁹.

Business plans and sustainability – a summary of key actions

- Based on your strategy, develop a phased business plan covering projections for 3, 5 and 10 years;
- Identify predicted costs and planned expenditures, together with an indication of the financial year in which expenditure will occur;
- Consider whether any costs can be recouped, for example by charging for services;
- Undertake a cost/benefits analysis to help make the case for investment;
- Address sustainability issues and the associated long-term costs.

In terms of detail, break down your plans according to the sustainability issues applicable to individual components of the service over the medium- to long-term. Replacement, development and maintenance costs for equipment and facilities will be fairly obvious, but you should also account for changes in the level of service that buy-in from the research community could bring. How far, for example, will the role of the RDM services staff change once the service has become routine and embedded in the research process? What

aspects will become less important and which new areas of activity could become the focus of increased demand? Can you build toward a less labour-intensive centre with greater automation and devolution of processes? Where these questions can be answered in the short- to medium-term, include predicted costs; but do not omit them from longer-term plans, where they should be shown as indicative and reinforcing your ownership of the direction and scope of the service.

5.3 Guidance, training and support

Various levels of guidance and training should be provided to meet different audiences. Some generic support, for example via institutional RDM guidance webpages, can be complemented with more tailored support. One-to-one consultancy sessions, such as those offered at the University of Northampton, allow researchers to ask specific questions so they can adopt relevant approaches to creating, managing and sharing their data²⁰. A range of options is needed to engage different groups so think broadly about what will be provided.

Guidance and helpdesks

Basic guidance is required on all aspects of data management and several universities have produced websites that collate best practice and direct researchers to local support²¹. These tend to cover the whole research lifecycle from applying for funding, through creating and managing data, to long-term preservation and reuse. Guidance is typically pragmatic and gives basic advice such as how to structure, name and version data, control access, and identify relevant data centres. There are a number of excellent sources of best practice that you can use such as the UKDA guide on managing and sharing data²².

Such websites are usually put together by undertaking an internal review of existing support. Relevant content can be searched for using common RDM-related terms, such as IPR, data ownership, repository, storage, backup and research computing. Once preliminary material has been collated, some form of engagement with support services (e.g. via emails, workshops or interviews) can unearth additional content to include. When drafting the content, general guidance can be copied and customised from some of these existing websites which are often available under Creative Commons licences.

¹⁷KRDS/I2S2 Digital Preservation Benefit Analysis Tools Project website, URL: <http://beagrie.com/krds-i2s2.php>

¹⁸For example: Beagrie, N. & Pink, C. (2012) *Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners*. Charles Beagrie Ltd & University of Bath. Retrieved 12 March 2013, from <http://opus.bath.ac.uk/32509>

¹⁹Starting in 2013, the DCC will be convening events and producing guidance that explains funder rules and expectations.

²⁰Jones, S. (2012, October 31). *The value of a one-to-one*. Blog post. DCC. Retrieved 12 March 2013, from <http://www.dcc.ac.uk/blog/value-one-one>

²¹DCC. *List of RDM guidance websites produced by UK universities*, URL: <http://www.dcc.ac.uk/resources/policy-and-legal/rdm-guidance-webpages/rdm-guidance-webpages>

²²Van den Eynden, V., et al. (2011). *Managing and sharing data: best practice for researchers*. 3rd edition. Essex: UK Data Archive. Available online: <http://data-archive.ac.uk/medial/2894/managing-sharing.pdf>

Helpdesk services may also be required. Several universities have set up a generic email address to filter RDM queries. Existing helpdesk systems should be used where possible and schema of typical questions can be developed to assist in routing enquiries. It is also worth bearing in mind that some studies have identified a desire for named contacts in preference to generic helpdesks²³. Making contact details more visible, or introducing support staff on training courses to raise awareness of the help they can provide may assist researchers to get the most out of the RDM services on offer.

Training for different audiences

Data management training tends to fall into two main categories:

1. courses for researchers, often with a discipline-specific focus or aimed at postgraduates;
2. Continuing Professional Development (CPD) to reskill support staff such as librarians.

Training for researchers is best developed in partnership with academic staff or disciplinary data experts (such as those employed by data centres) to ensure that content is both relevant and meaningful. Such partnerships can also enhance the integration of good practice into the research environment through the incorporation of RDM messages in existing induction and training programmes. Two examples of how this can be achieved²⁴ are the Open Exeter project which has involved PhD students in the development of their training programme and the DataTrain initiative which brought together researchers and data centre staff.

A large body of training materials is already available for repurposing, as outlined in the table below.

Title	Description	Target audience	Link
UKDA training materials	Slides and exercises for a course covering all aspects of the data lifecycle.	Researchers	http://www.data-archive.ac.uk/create-manage/training-resources
Research Data MANTRA	An online RDM training course with quizzes, videos and software tutorials.	Researchers	http://datalib.edina.ac.uk/mantra
CAiRO	A online RDM module for creative arts researchers.	Researchers	http://www.projectcairo.org/module/unit1-0.html
DataTrain	Slides and training materials for archaeology and social anthropology researchers.	Researchers	http://www.lib.cam.ac.uk/dataman/datatrain/datatrainintro.html
DATUM for health	Slides with speaker notes and audio recordings for health science researchers.	Researchers	http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/health/materials/?view=Standard
Introducing research data	A 27-page handbook with case studies and associated presentation.	Researchers	http://eprints.soton.ac.uk/338816
Leeds RoaDMaP materials	Presentations, handbook and feedback from courses aimed at engineering researchers, social scientists and research support staff.	Researchers and support staff	http://library.leeds.ac.uk/roadmap-project-outputs

²³Ward, C., et al. (2010). *Incremental: scoping study and pilot implementation plan. (ref. p20)*. Retrieved 12 March 2013, from http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf

²⁴Open Exeter project website, URL: <http://as.exeter.ac.uk/library/resources/openaccess/openexeter> and DataTrain project website, URL: <http://www.lib.cam.ac.uk/preservation/datatrain>

Title	Description	Target audience	Link
TraD – Training for Data Management at UEL	Online modules (forthcoming), slides and exercises for a variety of audiences.	Researchers and Librarians	http://www.uel.ac.uk/trad
DCC roadshows	Case studies, presentations and exercises aimed at support staff establishing RDM services.	Research support staff	http://www.dcc.ac.uk/events/data-management-roadshows
RDMRose	8 sessions with presentations, case studies and activity sheets.	Librarians	http://rdmrose.group.shef.ac.uk/?page_id=10
Essentials 4 Data Support	Introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data.	Data supporters	http://datasupport.researchdata.nl/en/
DIY Research Data Management Training Kit for Librarians	Presentations with accompanying audio and exercises demonstrating how the MANTRA module can be reused for academic liaison librarians.	Librarians	http://datalib.edina.ac.uk/mantra/libtraining.html

Given the availability of proven training materials, the effort required to put on courses for researchers is often focused on repurposing content and embedding provision. A trend observed in 2012 was to add RDM training into existing core curricula, and an approach being trialled by the University of Bath is to use Doctoral Training Centres as catalysts for change. By training each year's cohort of interdisciplinary students they hope to connect with a range of academic staff and students across the institution and influence the overall culture within the graduate school. Other institutions, such as the University of Northumbria, have secured the insertion of data management into core PhD skills courses. Targeting researchers early on in their careers can be useful, since good RDM practice can become embedded before less rigorous habits are formed. Also, if you can build training into existing programmes there will be an even greater chance of sustainability which makes it a better route to follow than offering one-off courses.

Similar lessons can be applied when developing training for support staff. Again there are a number of existing courses such as Data Intelligence for Librarians as well as a clutch of DCC courses that can easily be adapted for local use. Do bear in mind that different messages may be needed for librarians, information technology services staff and staff from the research office, who will

each approach the subject with a different professional heritage. Content should be tailored to focus on each group's particular skills and needs. It is also useful to add details explaining the broader institutional context so staff know what is provided by others in the organisation. In that way training is a good means of raising awareness about what is already in place, since this is often lacking.

Consultancy services

In some cases, researchers may require more hands-on, tailored support. It may be that they are just looking for a 'sounding board' to check the appropriateness of their data management techniques. Several universities, particularly in the USA, provide short consultations in this vein to help researchers to develop data management plans. A brief discussion about the research and proposed methods to create, manage and share data is likely to uncover areas where further guidance or pointers to useful services would be useful. It also allows the institution to check the appropriateness of plans and that relevant support and infrastructure is included in costing.

More in-depth support may also be needed during research projects, particularly in terms of delivering technical aspects such as database design. In some

²⁵Cope, J. (2011, December 15). Doctoral Training Centres as catalysts for research data management. Blog post. Research360 project, University of Bath. Retrieved 12 March 2013, from <http://blogs.bath.ac.uk/research360/2011/12/doctoral-training-centres-as-catalysts>

institutions, researchers have requested that IT services provide a dedicated pool of research support staff who could be costed into grant proposals to provide whatever technical support is needed. A small-scale example of this is provided in the College of Arts at the University of Glasgow where several members of technical staff are available to consult during bid development and are frequently costed in as technical partners²⁶.

Guidance, training and support – a summary of key actions

- Collate details of existing institutional support to provide a single, joined-up RDM guidance website;
- Coordinate the provision of support, either through a central helpdesk or well-signposted contacts;
- Provide RDM training to support researchers and reskill support staff;
- Offer more in-depth consultancy services as appropriate.

5.4 Data management planning

The need for data management planning services

Several major research funders in the UK require the inclusion of a data management plan (DMP) as an integral component of research grant applications²⁷. Whilst each funder may have a different emphasis, they generally require a basic outline of the data to be collected and an explanation of how it will be managed, shared and preserved, justifying any restrictions that need to be applied. The DMP is an opportunity to demonstrate awareness of best practice and reassure funders that data will be managed in line with their policies.

Universities are also asking researchers to create DMPs. Of the initial 15 institutional RDM policies released by UK universities by February 2013, 13 require or encourage the creation of DMPs as the procedure by which researchers should comply with their expectations. DMPs can be a very useful tool for institutions: they provide an opportunity to gather details on expected data volumes to assist in capacity planning; help to identify datasets to be recorded in institutional catalogues; and allow early engagement to validate the appropriateness of proposed approaches.

However, compliance with the policies of funders and institutions, whilst important, is not the sole reason

to encourage data management planning. There are many benefits for researchers to gain from the process. Planning saves time and effort. It enables researchers to make informed decisions, bearing in mind the wider context and consequences of different options, in order to anticipate and avoid problems. By considering what data will be created and how, researchers can check that they have the necessary support in place. The DMP process is ultimately most useful to researchers as it makes the research process easier.

Researchers are familiar with the grant submission process, so many may assume that they will be confident in developing DMPs. On the contrary, data management often presents a new set of challenges for researchers. They may not be aware of the support and services that are available within their institution and are often unsure of existing best practice and standards that they can adopt. The provision of guidance, tools and support services is crucial to help researchers create sound DMPs.

There are many approaches that universities are adopting to support the creation of DMPs, including:

- creating templates, guidance and libraries of example DMPs;
- provision of tools such as DMP Online, often customised with institutional guidance;
- offering training and advisory support such as consultancy services.

DMP templates and guidance

Institutions should provide templates or guidance on what should be included in plans, particularly when the institution requires a DMP. Various universities give an overview in their policies. The University of Hertfordshire, for example, provides a data planning checklist which lists seven themes that should be covered and a number of useful questions as pointers of what to address²⁸. Others have developed templates for specific audiences. The research360 project at the University of Bath has developed a very popular template for postgraduates²⁹. And there are several useful examples from the USA, such as the *What's your data plan?* guide from the University of Wisconsin-Madison³⁰. The DCC's *Checklist for a Data Management Plan*³¹ may also prove useful. This provides a very comprehensive list of questions to be considered across ten headings. It also contains a series of guidance notes designed to assist in the completion of a DMP, often accompanied by hyperlinks to useful and authoritative resources.

²⁶University of Glasgow. *Technical and data management support in the College of Arts webpage*, URL: <http://www.gla.ac.uk/services/datamanagement/whocanhelp/resourcedevelopmentofficers>

²⁷DCC. *Funders' data plan requirements*, URL: <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

²⁸University of Hertfordshire. (2011). *University guide to research data management. Policy appendix*. Retrieved 12 March 2013, from <http://sitem.herts.ac.uk/secreg/upr/pdf/IM12-apx%20III-University%20Guide%20to%20Research%20Data%20Management-v03%201.pdf>

²⁹Research360 project. (2012). *Postgraduate Data Management Plan template*. University of Bath. Retrieved 12 March 2013, from <http://blogs.bath.ac.uk/research360/2012/03/postgraduate-dmp-template-first-draft>

³⁰University of Wisconsin-Madison. *What's your data plan?* Retrieved 12 March 2013, from <http://researchdata.wisc.edu/make-a-plan/data-plans>

³¹Donnelly, M. & Jones, S. (2011, March 17). *Checklist for a Data Management Plan*. DCC. Retrieved 12 March 2013, from http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC_Checklist_DMP_v3.pdf

Guidance on appropriate methods should also be provided. Indeed, institutions may wish to prescribe a handful of recommended approaches for certain areas of work, such as storage and backup. Researchers are often unfamiliar with the support and services available to them, so providing associated guidance that raises awareness and provides links to support is invaluable. Requests for worked examples are also prevalent. ICPSR, a social science data archive in the USA, provides a *Framework for Creating a Data Management Plan*³². This is a great example of a template with basic guidance and worked examples. Some institutions have also considered compiling libraries of successful data management plans for researchers to learn from and reuse.

Data management planning tools

The DCC provides a web-based tool for creating, maintaining and exporting DMPs called DMPonline³³. DMPTool³⁴ provides a similar service in the USA. Both tools help researchers to write data management plans to meet funder requirements. DMPonline is organised according to which funder is the target of the grant submission and can be structured according to what stage has been reached (for example, application stage or in-project), making it a highly-focused tool.

As many universities in the UK also now require DMPs, DMPonline also allows institutions to create templates so plans can be created according to their requirements. A key benefit of the tool is how it can be customised by institutions. Universities can provide their own templates and guidance to ensure researchers are aware of local support, and can even brand the tool so it is seen as an institutional service³⁵. DMPonline offers a number of popular features, such as the sharing of plans to enable research teams and support staff to collaboratively develop their approach. The DCC is also revising the tool in light of user feedback and anticipates additional features will be available by summer 2013³⁶.

Training and consultancy

You may find that training on data management planning and more in-depth consultancy services are needed in addition to any guidance and tools you provide. Examples of training and models for consultancy are provided in section 5.3. You could also consider examples from overseas, for example the University of Virginia which has repurposed some of its library staff into a Scientific Data Consulting team³⁷. Anecdotal evidence suggests that consultancy services may be required more by researchers in the Arts and Humanities than the hard sciences.

Who is responsible for planning?

Data management planning is a collaborative endeavour. It must of course involve members of the research community, since a DMP establishes for them the manner in which they will capture or generate data. They will need to refer to it during the research lifecycle when the plan not only provides a guiding framework but may itself need to be amended to reflect unforeseen outcomes from the research. Creation of a DMP will also require professional support from the data librarians or repository staff responsible for the long term preservation and management of data, including the transmission of datasets to offsite centres. Input may also be required from members of the research support office, who will be responsible for issues of policy, finance and governance.

With such a diversity of interests maintaining effective communication will be vital. Perhaps some form of administrative or coordinating responsibility for data management planning will be useful, possibly even the central management of draft and completed plans. Together with training in the use of planning tools, these aspects will underpin any successful service.

In time, data management planning will be a skill acquired by all new postgraduate researchers. Until then, this relatively new practice will require support and advice. HEIs should again think broadly in terms of the range of support services they offer.

Data management planning – a summary of key actions

- If you have a requirement for DMPs, develop a template or guidelines on what to include;
- Provide guidance, approved options and examples to help researchers complete DMPs;
- Consider the use of tools such as DMPonline, customising these to your institution;
- Build data management planning into existing training and consultancy programmes as required.

³²ICPSR. (n.d.). *Framework for Creating a Data Management Plan*. Retrieved 12 March 2013, from <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html>

³³DMPonline, URL: <https://dmponline.dcc.ac.uk>

³⁴DMPTool, URL: <https://dmp.cdlib.org>

³⁵Donnelly, M. (2012, May 3). *Bringing it all back home: tailoring DMP Online for your institution*. Blog post. DCC. Retrieved 12 March 2013, from <http://www.dcc.ac.uk/blog/tailoring-dmp-online-for-your-institution>

³⁶Ashley, K. (2013, January 11). *Future plans for DMP Online*. Blog post. DCC. Retrieved 12 March 2013, from <http://www.dcc.ac.uk/news/future-plans-dmponline>

³⁷University of Virginia Library. *Scientific Data Consulting website*, URL: <http://www2.lib.virginia.edu/brown/data>

5.5 Managing active data

Two primary concerns when delivering services to support the management of data during the active phase of research are the provision of:

1. sufficient volumes of research data storage to ensure broad uptake and use;
2. relevant applications that offer the flexibility and functionality required by researchers to store, access and share their data during research collaborations.

Research data storage

If you are not aware of the quantity of research data being created within your institution, or do not know where this is held and backed up, a preliminary study to understand the scale of the problem is worthwhile. Requirements-gathering exercises have uncovered numerous incidences of hand-crafted approaches to research data storage, often referred to informally as DUDs – Data centres Under Desks. Faced with substantial charges from IT services for additional managed storage, research groups have opted to buy cheap storage and to run their own systems. While the upfront costs may be only a fraction of those quoted by central services, the risks of data loss and security breaches are significantly higher, potentially leading to far greater costs in the long term.

In response to this challenge, many universities are now providing much greater capacities of research data storage free of charge. A strong business case is crucial to securing the additional investment to allow this provision but, once funding has been committed, there are a number of storage options that can be pursued. Some universities are utilising their High Performance Computing (HPC) facilities while others are extending the capacity of existing filestores or exploring secure cloud storage options. Regardless of which route is taken, engagement with end users is critical to ensure that the proposed option will meet their needs.

Procedures also need to be developed to allocate and manage the storage. The model developed by the data.bris³⁸ project is of use here. Researchers are required to sign up as a data steward to be allocated 5 terabytes (TB) of storage, and are then responsible for controlling who has access and how long the data should be kept. Above 5TB the cost of storage is priced on a ‘Pay Once Store Forever’ basis, where ‘forever’ is defined as 20 years. Principal Investigators

anticipating a need for more than 5TB of disk storage are advised to include a request for funding in their grant applications.

Cloud storage services

Cloud services may be considered as an option to reduce capital investment and the need for expert skills to establish services in-house. However, researchers have expressed concern about the potential loss of control over their data and, whilst security generally matches or exceeds that provided by in-house systems, maintaining data security does become more challenging when data are distributed globally over a large number of devices that are being shared by a diverse community of unrelated users. The selection of cloud services for data storage is therefore a matter of judging the balance of risk to your data that may be acceptable when compared to the advantages from cost containment and ease of access.

The UK’s Janet Brokerage³⁹ is one route we recommend to the use of trusted services since it aims to strike relationships between the HE sector and suppliers or between groups within the sector with the objective of developing a community cloud of dynamically available resources. It has a major framework in place offering eight suppliers and one framework supplier already partnering with one institution. Work is also in progress towards a sector-wide deal with Microsoft and others are anticipated for Amazon Web Services, Google, Dropbox and Microsoft Azure. If you are considering the inclusion of cloud services within your data storage strategy it is worth bearing in mind that the Janet Brokerage works closely with representative bodies such as RLUK, UCISA and UUK as well as the Research Councils. The Janet brokerage offers advice and support to institutions moving to cloud computing and data centre services through white papers, online briefings and events.

Decisions about data storage solutions will encompass not only the selection of preferred platforms, tools and the extent to which third-party services will be used. They will also be influenced by the planned internal reach and sophistication of the institutional research data management service. This in turn will depend upon the prevailing research culture; for example, how devolved research support is and how much the research community looks for support to the centre.

³⁸Data.bris project, URL: <http://data.blogs.llrt.org>

³⁹Janet brokerage website, URL: <https://www.ja.net/products-services/janet-brokerage>

Academic 'dropbox' services

A common area of interest within current RDM projects is the provision of dropbox-like services. Researchers regularly use Dropbox as it allows them to access and work on their data from multiple devices, automatically syncing back to a central copy. Data is backed up and files can be shared publicly or privately with other users. This is often far easier for remote working and collaboration than operating via centrally-managed networked storage, particularly when collaborators are based in other HEIs and organisations.

Due to the perceived security and potential legal risks from using third-party services, many universities have been investigating options for running services that can be kept firmly under their own control. The Universities of Lincoln and Edinburgh have both piloted OwnCloud, an open source alternative to Dropbox⁴⁰, although neither found it sufficiently developed at present for non-technical users to implement it without ongoing support. Elsewhere, the University of Oxford has developed DataStage, a tool to allow research groups to share their data and collaborate⁴¹.

Given the demand for dropbox-like services, it has been proposed that the Janet Brokerage should attempt to negotiate a deal with Dropbox on behalf of the HE sector. Institutions could then use the service with the assurance of appropriate legal safeguards, in the knowledge that the terms and conditions comply with funder expectations on the jurisdiction in which UK research data may be stored.

RDM platforms

Complete research data management systems may also be developed in-house. At the University of Manchester the MaDAM project⁴² developed a prototype system in collaboration with biomedical researchers. Several other examples of disciplinary systems are also available, such as OMERO⁴³ and BRISKit⁴⁴. OMERO is a tool for microscope images. It handles hundreds of image formats, allowing researchers to gather all their images in a secure central repository where they can view, organise, analyse and share the data from anywhere via internet access. The BRISKit project meanwhile is designing a national shared service brokered by Janet to host, implement and deploy biomedical research database applications that support the management and integration of tissue samples with clinical data and electronic patient records.

The RDM platforms section of the Monash University case study that complements this guide is particularly instructive. Monash has adopted a mantra to “adopt; adapt; develop”. If a research community already has a solution (or there is an emerging one), they adopt this and, where necessary, adapt it to suit the needs of researchers at Monash. Only as a last resort will an entirely new solution be developed. Developing a new product may be expensive, costly to support, and could split researchers from their community. This approach acknowledges that researchers' loyalty is often stronger to their discipline than their institution, and that close engagement with researchers is critical when designing RDM systems to ensure their applicability and uptake.

The common requirements that seem to be emerging are for a globally-accessible cross-platform filestore that provides all collaborators with access to the data, regardless of where they are based. Options for backup and synchronisation of data on mobile devices could also be considered. And provision for routine backup, long-term archiving and data sharing should always be addressed. To ensure the uptake of data management systems, appropriate volumes of storage need to be provided and the systems need also to be sufficiently flexible, most notably in terms of access, to fit with the gamut of researchers' working practices.

Managing active data – a summary of key actions

- Review data holdings and RDM practices to see if the current infrastructure and systems are sufficient;
- Where appropriate, make a case for investment to provide additional research data storage;
- Develop procedures for the allocation and management of research data storage;
- Provide flexible RDM systems to support the creation, management and sharing of data that meet a diverse range of research contexts and needs.

⁴⁰Winn, J. (2012, August 6). OwnCloud: An 'academic dropbox'? Blog post. *Orbital project*, University of Lincoln. Retrieved 12 March 2013, from <http://orbital.blogs.lincoln.ac.uk/2012/08/06/owncloud-an-academic-dropbox>

⁴¹DataFlow. DataStage webpage, URL: <http://www.dataflow.ox.ac.uk/index.php/about/about-datastage>

⁴²MaDAM project, URL: <http://www.library.manchester.ac.uk/aboutus/projects/madam>

⁴³OMERO website, URL: <http://www.openmicroscopy.org/site/products/omero>

⁴⁴BRISKit website, URL: <https://www.briskit.le.ac.uk>

5.6 Data selection and handover

Why select data for retention?

The UK Research Councils expect data with 'acknowledged long-term value' to be preserved and remain accessible and usable for future research⁴⁵. Moreover, they require that data management activities are both efficient and cost-effective in the use of public funds. If institutions are to ensure appropriate use of public funds, a selection process is essential to prioritise data for long-term curation.

While there is a cost to selecting data for retention, this is minimal in comparison with the cost of keeping everything, especially when that means making it accessible online. It is true that storage costs have fallen at a rate of 30% per annum since the 1980s but this is unlikely to continue according to preservation expert David Rosenthal⁴⁶. Factors working against the trend include the massive increase in digital data volumes. According to estimates made in 2012 by market analysts IDC, the 'digital universe' will be 50 times larger in 2020 than it was in 2010⁴⁷. While demand for storage is spiralling upwards, the costs of virtual storage are not falling at the rate that disk storage historically has, making it difficult to predict long term storage costs⁴⁸.

A selection process may take time to set up and operate, but that should pay off in terms of the ability to forecast preservation and storage costs. Selection also counters risks to the institution. These include the reputational damage from exposing dirty, confidential or undocumented data that has been retained long after the researchers who created it have left. A first step in establishing a selection process is to determine broad categories of data that are aligned with the institution's mission. First on the list will be data that you are legally obliged to retain, for example for contractual or regulatory reasons. Policy obligations may also be high on the list. Publishers and learned societies increasingly require authors to archive the data supporting their results in an appropriate public archive. The emphasis of RCUK policy is on retaining the evidence base to support the scholarly record: the data which directly underpin publications or other research outputs.

Encouraging researchers to select and deposit data

It is standard practice in certain disciplines to archive where data are a recognised research output and

subject repositories are well-established. In others, data sharing may be novel and threatening, and even the term 'data' is seen as problematic. Policies or guidelines that expect researchers to hand over their data may feel threatening, especially if the process involves taking decisions on what is worth keeping. This makes establishing guidelines, processes and good practice for data selection and deposit one of the more challenging aspects of an RDM service. Obtaining buy-in from researchers is essential. This is likely to call for a combination of:

- high level guidance from central services;
- advocacy and guidance at a department or research group level;
- deposit tools to make it as easy as possible for researchers to hand over data;
- deposit agreements that establish what will happen with the data and engender trust between researchers and the institution.

High-level guidance

Clear guidelines are needed to set parameters similar to those contained in a Data Centre or Archive Collections Policy⁴⁹. The guidelines should be clear about which data fall within the institutional service's remit, what kinds of data will be accepted and which are the priority areas. Also consider what range of outcomes is envisaged as different levels of service may be applied according to the potential reuse value of the data. Some examples of these options include the following:

- provide recommendations to researchers on local archiving or disposal options;
- catalogue the data and provide an assured level of destruction, e.g. for confidential data;
- catalogue the data and deposit it for preservation in a recognised subject repository or cross-institutional data store;
- catalogue and preserve the data in the institutional data repository for sharing under agreed conditions.

You may wish to apply different levels of curation based on the value of the data you accept. At the most basic end of the spectrum, institutions should record metadata to aid discoverability. In cases

⁴⁵RCUK. (2011). *Common Principles on Data Policy*. Retrieved 12 March 2013, from <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx> Funders' data plan requirements, URL: <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

⁴⁶Rosenthal, D. (2012, May 14). *Let's just keep everything forever in the cloud*. Blog post. Retrieved 12 March 2013, from <http://blog.dshr.org/2012/05/lets-just-keep-everything-forever-in.html>

⁴⁷IDC. (2012). *The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East*. Retrieved 12 March 2013, from <http://www.emc.com/leadership/digital-universe/iview/executive-summary-a-universe-of.htm>

⁴⁸Rosenthal, D. & Vargas, D. (2013) *Distributed digital preservation in the cloud*. Paper given at 8th International Digital Curation Conference, Amsterdam January 2013. Retrieved 12 March 2013, from <http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf>

⁴⁹See for example: *Archaeology Data Service*. (2007, March 21). *Collections policy*. 4th edition. Retrieved 12 March 2013, from <http://archaeologydataservice.ac.uk/advice/collectionsPolicy>

where data can't be shared but must be retained for a certain period, you may opt to perform basic archiving (i.e. storing, backing up and performing periodic integrity checks). When data are to be preserved and shared, increasing degrees of processing can be applied depending on the condition of the data and documentation and the anticipated level of reuse. Beyond the 'relevance to mission', a central support service can advise on many key considerations including the likely scale of preservation costs and technical criteria such as the use of appropriate formats.

Advocacy and departmental guidance

Selection is also about judging whether the data are useful in scientific or scholarly terms, and what contextual information, software or hardware details are needed to make them so. The decisions here will lie with the original researchers, who should be prepared to justify their decisions based on recognised best practice in their discipline.

Relevant data centres will be a useful source of advice on this. For example the NERC data value checklist⁵⁰ supports researchers both at the pre-award stage of data management planning and the pre-deposit stage of identifying which data should be deposited with the NERC Environmental Data Centres. It is a weighted list of criteria, phrased as questions to help identify which data are of long-term value. The UK Data Service is developing a data appraisal kitemark⁵⁰. This similarly uses a set of scales to determine which data should be prioritised for selection.

Departmental or research group guidance should consider what is meaningful to preserve, and give examples of potential uses for raw data, data processed to interim stages of analysis or conditioning, or final outputs. Depending on the discipline, reproducing the study results may require all or some of these categories. In some disciplines it will be more important to scrutinise and appreciate the processes followed than to reproduce the results. Other reuse cases could include the production of teaching and learning materials.

A central support service can provide input and advice. For example library colleagues may be qualified to advise on:

- potential for reusing the data in other disciplines;
- generic criteria that can help to prioritise effort for getting data into a reusable state;

- relevant guidance from data centres, learned societies and publishers;
- how data may be used in e-learning and outreach.

We describe generic scientific and technical criteria under 'Establishing criteria for selection decisions' below.

Researchers are likely to make decisions about what to share when they have reached the stage of planning final outputs, whether that means writing up for publication or preparing a software release or an exhibition. It makes sense to use such 'organisational moments'⁵² towards the end of projects to provide information on short and long-term benefits, and the services available to help achieve them.

Deposit tools

Jisc has supported the development of various tools to facilitate repository deposit. The DataFlow and SWORD-ARM⁵³ projects, for example, both make use of the SWORD2 protocol to ease the deposit process. Several deposit scenarios are outlined on the SWORD blog⁵⁴. Other tools such as DepositMOre⁵⁵ and DataUp⁵⁶ make use of widespread software (Microsoft Word and Excel respectively) by embedding options to deposit directly.

You should also consider how to ease the creation of metadata, which is often the hurdle that puts researchers off deposit. Tools such as DataFlow and DataUp make it easy to add basic data descriptors in the environments used for managing active data and automatically extract the metadata from there. Involving and consulting researchers in defining deposit workflows that fit their practices should also pay off in achieving better uptake.

Deposit agreements

Researchers need to feel in control when they hand over the fruits of their labour and should be able to rely on getting back what they put in. They can reasonably also expect to benefit from some form of added value. A deposit agreement will set out terms and conditions that communicate the responsibilities of depositor and service provider. The agreement should give the repository rights to manipulate the data, as preservation may require migration to new formats. It should also allow the repository to reserve the right to withdraw the data for legal or other reasons. For an example deposit agreement see that by the University of Edinburgh⁵⁷.

⁵⁰NERC. (2012). Data value checklist. Retrieved 12 March 2013, from <http://www.nerc.ac.uk/research/sites/data/dmp.asp>

⁵¹A draft form of the appraisal kitemark can be seen on slide 12 of the presentation Data Appraisal at the UK Data Archive, retrieved 12 March 2013, from <http://researchdataessex.posterous.com/jisc-mrd-programme-workshop-2425-oct-rde-post>

⁵²See for example: Garrett, L. et al. (2012). JISC funded Kaptur project environmental assessment report. Project Report. Visual Arts Data Service (VADS). Available online: <http://www.research.uct.ac.uk/1054>

⁵³SWORD-ARM project, URL: <http://archaeologydataservice.ac.uk/research/swordarm>

⁵⁴SWORD blog, URL: <http://swordapp.org/2012/07/data-deposit-scenarios>

⁵⁵DepositMOre blog, URL: <http://blog.soton.ac.uk/depositmo>

⁵⁶DataUp tool, URL: <http://dataup.cdlib.org>

⁵⁷University of Edinburgh. DataShare depositor agreement, URL:

<http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/data-repository/depositor-agreement>

Establishing criteria for selection decisions

You should establish criteria to guide selection decisions. The DCC's *How to Select and Appraise Research Data for Curation*⁵⁸ proposes seven criteria as outlined below:

- 1. Relevance to mission:** the resource content fits any priorities stated in the institution's mission, or funding body policy including any legal requirement to retain the data beyond its immediate use.
- 2. Scientific or historical value:** is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
- 3. Uniqueness:** the extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere.
- 4. Potential for redistribution:** the reliability, integrity, and usability of the data files may be determined; these are received in formats that meet designated technical criteria; and Intellectual Property or human subjects issues are addressed.
- 5. Non-replicability:** it would not be feasible to replicate the data/resource or doing so would not be financially viable.
- 6. Economic case:** costs may be estimated for managing and preserving the resource, and are justifiable when assessed against evidence of potential future benefits; funding has been secured where appropriate.
- 7. Full documentation:** the information necessary to facilitate future discovery, access, and reuse is comprehensive and correct; including metadata on the resource's provenance and the context of its creation and use.

Institutional guidelines based on these criteria, coordinated with local guidelines that adapt them to specific disciplinary contexts, will help to focus resources on data that have real potential for impact and reuse. Services such as access management, storage and preservation need to be prioritised according to value and user demand, a principle familiar to any library user who has borrowed from a short loan collection or sought a rare item from an archive. A similar approach needs to be adopted for

data; much of what is retained need not be online or even onsite, and the level of additional care it requires will vary.

Data selection and handover – a summary of key actions

- Identify which data fall under your remit and establish criteria to guide decisions on what to keep;
- If establishing a repository, develop deposit agreements and high-level guidance;
- Develop or use existing deposit tools to ease the process of handover;
- Advocate the benefits of data deposit to encourage uptake;
- Support research groups to develop guidance and offer input to decisions;
- Establish criteria to apply differing levels of curation depending on the nature of the data.

5.7 Data repositories

Once you have established which data to keep, an obvious question to address is how they will be preserved and shared. There are arguments for taking a DIY approach, and also for using external services, but it is most likely that institutions will adopt some combination of these.

Possibly the first consideration will be cost, including the price to be paid for maintaining a data repository and the provision of expert staff to run it. It is generally accepted that costs are heavily weighted around the point when data are first deposited (or ingested, from the repository perspective) as that's where most intensive staff activity takes place. However further costs will arise depending on how much preservation activity will be required over the longer term and how long that period is.

The retention period and preservation costs beg consideration of factors familiar to the records manager. There are a number of statutory and business requirements affecting particular types of data plus emerging mandates imposed by the major funders, some of whom dictate how long research data should remain discoverable and accessible. Expectations range from three years to 'in perpetuity',

⁵⁸Whyte, A. & Wilson, A. (2010). 'How to Select & Appraise Research Data for Curation'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

with most funders expecting data to be preserved for ten years or more.

The availability of existing services such as subject-specific repositories and designated data centres provided by research funders should also be factored into your considerations. It is arguably better for research data to reside with other data from that discipline, so you may choose to recommend that researchers use existing services where they are in place. However, the likelihood is that most research data produced in your institution will not be served by an existing data service, so a hybrid approach is probably needed.

Finding a solution requires the consideration of three basic options:

- development and maintenance of an institutional data repository;
- liaison with external research data repositories;
- signposting of relevant services for researchers.

Institutional data repositories

Institutional repositories have typically been created to store research publications rather than data, but their technical infrastructure can be extended to enable the curation of data without the development or purchase of an entirely new software platform. If you plan to use your existing repository to manage data we recommend that you seek advice from another institution using the same repository software. For example, if you have an ePrints platform you might want to investigate the DataPool⁵⁹ project, which explains how the University of Southampton is exploring the development of a full research data management service using ePrints and SharePoint software. The Universities of Cambridge and Edinburgh both run DSpace repositories for research data⁶⁰ and several institutions are exploring the use of CKAN to catalogue and store data⁶¹.

Where repositories are being used to support submissions to the Research Excellence Framework, options exist to align the research data store with commercial research information management systems such as Atira PURE and Symplectic Elements. One way of discovering how others have tackled the installation or extension of repositories would be to join a group such as UKCoRR⁶², (the United Kingdom Council of Research Repositories), an independent body for repository managers, administrators and staff in the UK. Jisc also provides support to repositories on particular aspects of

infrastructure delivery; for example the Jisc-funded RIOXX project⁶³ is engaged in developing guidelines for repository metadata. The key message is not to struggle on your own but seek help from others already engaged in repository development or operation.

If you do not already have a data repository, inevitably there will be a significant capital cost for setting up a secure facility. A less obvious but more persistent revenue cost will derive from the necessary human infrastructure and any institution deciding to store and curate its own data will need first to understand who will be providing the repository service, what essential skills are currently available and which existing staff can be retrained into a data management role, as well as what new posts must be created, funded and staff recruited. Frequently the repository is offered as a new library service and it will be easy to identify sources of expertise from amongst existing staff. Otherwise, a good start would be to run a CARDIO assessment in order to build a consensus of requirements and expectations across the range of principal stakeholders.

Decisions on preferred options for archival data storage cannot be taken in isolation from other defining factors. It is important to assess which services and tools you will need to address the full range of actions set out in the DCC's Curation Lifecycle. The DCC tools and services catalogue⁶⁴ provides a wide range of guidance on curation and preservation actions. The DCC Lifecycle model is derived partly from the Open Archival Information System (OAIS) standard⁶⁵, which gives a more detailed schema for developing preservation services and describes how these may interact. Using the Lifecycle model and OAIS standard can help you maximise the functionality of an in-house data store and to plan how services will be embedded alongside services for managing research information and publications.

External research data repositories

There are hundreds of existing data repositories worldwide, some of which may already be used by researchers in your institution. Many of these are discipline-specific or community-based, some are linked to publishers, others have grown out of small start-ups, for example figshare, while yet more are large international initiatives such as the World Data Centre. A useful list of research data repositories from around the world is available via Databib⁶⁶. This shows which subject areas are supported by each repository and outlines any restrictions on data access, licence agreements and the identifiers used.

⁵⁹DataPool project, URL: <http://datapool.soton.ac.uk/datapool>

⁶⁰DSpace@Cambridge, URL: <http://www.dspace.cam.ac.uk> and Edinburgh DataShare, URL: <http://datashare.is.ed.ac.uk>

⁶¹See for example: Winn, J. (2012, September 6). Choosing CKAN for research data management. Blog post. Orbital project, University of Lincoln. Retrieved 12 March 2013, from <http://orbital.blogs.lincoln.ac.uk/2012/09/06/choosing-ckan-for-research-data-management>

⁶²UKCoRR, URL: <http://ukcorr.org>

⁶³RIOXX project, URL: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/repositories/riox.aspx

⁶⁴DCC. Tools and services catalogue, URL: <http://www.dcc.ac.uk/resources/external/tools-services>

⁶⁵Higgins, S. (2007). Using OAIS for Curation. DCC Briefing Paper. Available online: <http://hdl.handle.net/1842/3354>

⁶⁶Databib website, URL: <http://databib.org>

The immediate benefit of using an established data repository is the access it provides to a ready-made and expert infrastructure, not only for storing data but also for enabling its discovery and delivery. While data centres help 'to set standards for data creation and metadata, they cannot provide the individual, tailored support that individual researchers need as they work on their individual projects'⁶⁷. However, some attempts are being made to liaise with researchers prior to data handover, particularly in the case of data centres sponsored by research funders.

Some funders directly support facilities to curate the data generated from the research they have sponsored. Although there is an onus on researchers to offer data for deposit, designated services such as the UK Data Archive⁶⁸ and NERC data centres⁶⁹ nonetheless apply strict criteria to determine whether data will be accepted. There may also be a cost. The Archaeology Data Service⁷⁰, for example, expects to recover the cost of archiving from the body funding the archaeological investigation, in the form of a one-off payment collected at the time of deposit⁷¹. Before deciding how to incorporate the data centres into any research data management strategy, it is essential to understand the rules of engagement for each of them. As a first step, consult the table in the DCC's overview of funders' data policies⁷², where we indicate which of them support data centre services.

Signposting relevant services for researchers

Researchers and data managers are likely to be well informed about available subject repositories in their discipline. If there are none, there is potential for the RDM service to advise on any that are in closely-related fields and may accept deposit. The service should at least point researchers to lists of suitable repositories. The directory of data repositories available from Databib⁷³ may be of use.

The choice of any external repository should take into account the need to ensure that data deposited overseas is in a jurisdiction offering equivalent legal protection as it would receive in the UK. The repository should also be covered for the possibility that it ceases operation, by arranging to hand over the data to somewhere that can sustain similar protection. You may wish to provide a list of approved repositories that your institution has confidence in and is willing to recommend to researchers.

When you are developing your strategy, consider how far you can go in attempting to fund and coordinate researchers' use of external services and whether they meet institutional criteria for data preservation and sharing. For instance, should data deposited in such a repository be tracked by the institutional data catalogue? How will researchers notify the institution of such deposits and will a core set of the data be maintained onsite? Clear procedures need to be established.

Data repositories – a summary of key actions

- Where required, develop and maintain an institutional data repository;
- Explore options to align your repository with existing systems, for example for Research Information Management;
- Decide what role external data repositories will play in your RDM strategy;
- Provide guidance and support to direct researchers to relevant services.

5.8 Data catalogues

Research organisations are expected to have a record of the research data they hold and to make this metadata available online to support data discoverability and reuse. The RCUK *Common Principles on Data Policy* state that:

To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data⁷⁴.

The EPSRC goes further by recommending the use of robust Digital Object Identifiers (DOIs) and specifying that the metadata is sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it⁷⁵.

These requirements leave institutions with a number of issues to consider:

- what metadata is needed to adequately record datasets?

⁶⁷Collins, E. (2012). *The national data centres*. In G. Pryor (Ed.), *Managing Research Data* (pp 151-172). London: Facet. ISBN: 978-1-85604-756-2

⁶⁸UK Data Archive, URL: <http://data-archive.ac.uk>

⁶⁹NERC data centres, URL: <http://www.nerc.ac.uk/research/sites/data>

⁷⁰Archaeology Data Service, URL: <http://archaeologydataservice.ac.uk>

⁷¹Archaeology Data Service. (2007, November). *Charging policy. Version 4*. Retrieved 12 March 2013, from <http://archaeologydataservice.ac.uk/advice/chargingPolicy>

⁷²DCC. *Overview of funders' data policies*, URL: <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

⁷³Databib website, URL: <http://databib.org>

⁷⁴RCUK. (2011). *Common Principles on Data Policy*. Retrieved 12 March 2013, from <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

⁷⁵EPSRC. (2011). *Policy Framework on Research Data*. Retrieved 12 March 2013, from <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>

- does any of this metadata already exist, and if so, where is it held?
- how should the metadata be captured?
- how will the metadata be exposed, and if possible collated nationally?

It is useful to consult guidance from DataCite, an organisation that aims to establish easier access to research data and persistent identification via DOIs. It offers a metadata schema⁷⁶ that lists core metadata properties chosen for the accurate and consistent identification of data for citation and retrieval purposes, along with recommended use instructions. By working with data centres, DataCite assigns persistent identifiers to datasets, and is developing an infrastructure to support simple and effective methods of data citation, discovery, and access. Useful content is available via the British Library which has run a series of five workshops to support HEIs⁷⁷.

A number of institutions are attempting to define the metadata required for datasets. At the University of Oxford, the DaMaRO⁷⁸ project has been developing DataFinder. A three-tier metadata approach is envisaged, comprising:

1. mandatory minimal metadata – a set of 12 fields extended from DataCite metadata;
2. mandatory administrative information – e.g. funder details and grant number;
3. optional, discipline-specific metadata to enable reuse.

The University of Essex is similarly adopting a three-tier approach. Their schema is based on DataCite, INSPIRE and DDI 2.1⁷⁹, a combination of generic metadata schema and specific standards for social science data. Another group of universities involved in the C4D project is exploring options to create an extension to the CERIF standard to describe datasets⁸⁰. The use of accepted standards is key. Moreover, to enhance discoverability it makes sense to comply with global metadata harvesting initiatives such as that provided by OpenAIRE⁸¹.

Although no dominant model has yet emerged, these early innovators can hopefully lessen the work for other institutions. Earlier research is also worthy of note, such as the UKOLN Scientific Data Application Profile Scoping Study⁸². This study assessed whether a single metadata application profile for research

data, or a small number thereof, would improve resource discovery or discovery-to-delivery in any useful or significant way.

When considering how metadata can be captured, opportunities to automatically harvest data from related systems to avoid re-entry should be explored wherever possible. Universities are considering different options of where to house the metadata: existing repositories or research information management systems could potentially be extended, or specialised metadata stores could be employed⁸³. Options for collating metadata at a national level, for example as is done by Research Data Australia⁸⁴, are also being explored by Jisc and the DCC. A number of pilot studies will be compared to draw out lessons for the wider community and inform the development of national data registry services.

The primary concerns for HEIs are to:

- collect the metadata in a seamless way, integrating systems wherever possible to avoid placing additional administrative burdens on researchers;
- ensure that standards are followed wherever possible to enable export into any national system as it develops.

As yet, there is not an accepted model to follow; research into data catalogues is very much at the preliminary, pilot stages. However, we encourage people to monitor developments and adopt emerging approaches rather than each institution defining its own standard.

Data catalogues – a summary of key actions

- Learning lessons from early innovators, define the metadata you need to record research datasets;
- Establish a system for capturing and displaying a record of research data holdings;
- Integrate systems where possible to benefit from options to harvest data and embed metadata creation into existing workflows;
- Expose metadata for inclusion in national catalogues or other relevant services.

⁷⁶DataCite. Metadata scheme, URL: <http://schema.datacite.org>

⁷⁷DataCite workshops website, URL: <http://www.bl.uk/aboutus/stratpolprog/digi/datasets/workshoparchive/archive.html>

⁷⁸DaMaRO project blog, URL: <http://blogs.oucs.ox.ac.uk/damaro>

⁷⁹Ensom, T. (2012, October 19). Repository beta and metadata profile released. Blog post. Research Data @Essex. Retrieved 12 March 2013, from <http://researchdataessex.posterous.com/repository-beta-metadata-profile-released>

⁸⁰C4D project blog, URL: <http://cerif4datasets.wordpress.com>

⁸¹OpenAIRE website, URL: <https://www.openaire.eu>

⁸²Scientific Data Application Profile Scoping Study, URL: <http://www.ukoln.ac.uk/projects/sdaps>

⁸³ANDS. Metadata stores solutions, URL: <http://www.ands.org.au/guides/metadata-stores-solutions.html>

⁸⁴ANDS. Research Data Australia website, URL: <http://researchdata.ands.org.au>

6. Getting started

There are many services that an institution may provide to support research data management. While these may be led by different groups or functions, there should be a coherent vision across the services. The integration of systems and workflows is also key; for example many are attempting to join up research information systems with repositories, or to support data management planning through the inclusion of flags in grant application systems. Coordination is crucial to raise awareness and embed good practice.

The key to getting started is to identify your strengths and weaknesses and to prioritise action. The DCC's CARDIO tool⁸⁵ can help you take stock of your current position. Service development can then be approached incrementally, addressing each component as a manageable task whilst keeping an eye on their wider coordination.

Support from colleagues and senior management is crucial to achieve your aims and ensure sustainable, well-embedded services. As we have already remarked, a useful initial step is to convene a steering group to develop the strategy and oversee progress. To be effective this should include a senior academic champion, researchers, and a broad range of service staff. The DCC's briefing paper on *Making the Case for Research Data Management* will help you find ways to engage different groups⁸⁶.

7. Acknowledgement

We thank DCC colleagues for their input, particularly Laura Molloy and Florence Kennedy for proofreading. We are also very grateful for comments from: the Iridium project (University of Newcastle); Managing Research Data: a pilot project in Health and Life Sciences (University of the West of England); UK Data Archive (University of Essex); Robin Rice (University of Edinburgh); and Simon Hodson (Jisc). Many of the examples in this guide come from the Jisc MRD community and DCC Institutional Engagement partners, so we are indebted to those involved in these programmes for sharing their lessons.

8. Bibliography and further sources

Many of the examples within this guide are drawn from work undertaken during two phases of the Jisc Managing Research Data (MRD) programme and a series of DCC Institutional Engagements. We encourage you to consult these programmes of work in detail:

DCC Institutional Engagements: <http://www.dcc.ac.uk/community/institutional-engagements>

Jisc Managing Research Data programme 2011-13: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

Jisc Managing Research Data programme 2009-2011: <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

Beitz, A., Dharmawardena, K. & Searle, S. (2012). *Monash University Research Data Management Strategy and Strategic Plan 2012-2015*. Retrieved 12 March 2013, from <https://confluence-vre.its.monash.edu.au/download/attachments/39752006/Monash+University+Research+Data+Management+Strategy-publicrelease.pdf?version=1&modificationDate=1334289180000>

Beagrie, N. & Pink, C. (2012) *Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners*. Charles Beagrie Ltd & University of Bath. Retrieved 12 March 2013, from <http://opus.bath.ac.uk/32509>

Collins, E. (2012). The national data centres. In G. Pryor (Ed.), *Managing Research Data* (pp 151-172). London: Facet. ISBN 978-1-85604-756-2

Ensom, T. (2012, October 19). *Repository beta and metadata profile released*. Blog post. Research Data @Essex. Retrieved 12 March 2013, from <http://researchdataessex.posterous.com/repository-beta-metadata-profile-released>

EPSRC. (2011). *Policy Framework on Research Data*. Retrieved 12 March 2013, from <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>

Garrett, L. et al. (2012). *JISC funded Kaptur project environmental assessment report*. Project Report. Visual Arts Data Service (VADS). Available online: <http://www.research.ucreative.ac.uk/1054>

⁸⁵CARDIO website, URL: <http://www.dcc.ac.uk/projects/cardio>

⁸⁶Whyte, A. & Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/briefing-papers>

Higgins, S. (2007). *Using OAIS for Curation*. DCC Briefing Paper. Available online: <http://hdl.handle.net/1842/3354>

ICPSR. (n.d.). *Framework for Creating a Data Management Plan*. Retrieved 12 March 2013, from <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html>

Jones, S., Ball, A. & Ekmekcioglu, Ç. (2008). The Data Audit Framework: a First Step in the Data Management Challenge. *International Journal of Digital Curation*, 3 (2): 112–120. doi:10.2218/ijdc.v3i2.62

Research360 project. (2012). *Postgraduate Data Management Plan template*. University of Bath. Retrieved 12 March 2013, from <http://blogs.bath.ac.uk/research360/2012/03/postgraduate-dmp-template-first-draft>

RCUK. (2011). *Common Principles on Data Policy*. Retrieved 12 March 2013, from <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

Rosenthal, D. & Vargas, D. (2013) *Distributed digital preservation in the cloud*. Paper given at 8th International Digital Curation Conference, Amsterdam January 2013. Retrieved 12 March 2013, from <http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf>

UKRIO. (2009). *Code of Practice for Research: Promoting good practice and preventing misconduct*. London: UK Research Integrity Office. Retrieved 12 March 2013, from <http://www.ukrio.org/what-we-do/code-of-practice-for-research>

University of Edinburgh. (2012). *Research Data Management (RDM) Roadmap August 2012 – January 2014*. Retrieved 12 March 2013, from <http://www.ed.ac.uk/schools-departments/information-services/about/strategy-planning/rdm-roadmap>

Van den Eynden, V., et al. (2011). *Managing and sharing data: best practice for researchers*. 3rd edition. Essex: UK Data Archive. Available online: <http://data-archive.ac.uk/media/2894/managingsharing.pdf>

Whyte, A. & Tedds, J. (2011). 'Making the Case for Research Data Management'. *DCC Briefing Papers*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/briefing-papers>

Whyte, A. & Wilson, A. (2010). 'How to Select & Appraise Research Data for Curation'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

Winn, J. (2012, August 6). *OwnCloud: An 'academic dropbox'*? Blog post. Orbital project, University of Lincoln. Retrieved 12 March 2013, from <http://orbital.blogs.lincoln.ac.uk/2012/08/06/owncloud-an-academic-dropbox>

Winn, J. (2012, September 6). *Choosing CKAN for research data management*. Blog post. Orbital project, University of Lincoln. Retrieved 12 March 2013, from <http://orbital.blogs.lincoln.ac.uk/2012/09/06/choosing-ckan-for-research-data-management>

Follow the DCC on Twitter:
@digitalcuration and #ukdcc

