Developing Research Data Management Capability: the View from a National Support Service

Sarah Jones
University of Glasgow
HATII, 11 University Gardens
Glasgow, G12 8QJ
+44 141 330 3549
sarah.jones@glasgow.ac.uk

Graham Pryor
University of Edinburgh
DCC, Appleton Tower
Edinburgh, EH8 9LE
+44 131 650 9985
graham.pryor@ed.ac.uk

Angus Whyte
University of Edinburgh
DCC, Appleton Tower
Edinburgh, EH8 9LE
+44 131 650 9986
a.whyte@ed.ac.uk

ABSTRACT
An increasing number of UK Higher Education Institutions (HEIs) are developing Research Data Management (RDM) support services. Their action reflects a changing technical, social and political environment, guided by principles set out in the Research Councils UK (RCUK) Common Principles on Data Policy. These reiterate expectations that publicly-funded research should be openly accessible, requiring that research data are effectively managed. The Engineering and Physical Sciences Research Council (EPSRC) policy framework is particularly significant, as it sets a timeframe for institutions to develop and implement a roadmap for research data management.

The UK Digital Curation Centre (DCC) is responding to such changes by supporting universities to develop their capacity and capability for research data management. This paper describes an 'institutional engagement' programme, identifying our approach, and providing examples of work undertaken with UK universities to develop and implement RDM services. We are working with twenty-one HEIs over an eighteen month period, across a range of institution types, with a balance in research strengths and geographic spread. The support provided varies based on needs, but may include advocacy and awareness raising, defining user requirements, policy development, piloting tools and training. Through this programme we will develop a service model for institutional support and a transferable RDM toolkit.

Categories and Subject Descriptors
E.0 [**Data General**].

General Terms
Management, Design, Security, Human Factors, Legal Aspects.

Keywords
Research Data Management, data sharing, university, higher education, infrastructure, research data policy, Data Management Plan, training, Digital Curation Centre, JISC.

1. INTRODUCTION
The desire among UK Higher Education Institutions (HEIs) to develop Research Data Management (RDM) roadmaps is driven by a range of factors. Developments in research data policy are a key influence, as are social and political demands for transparency. Controversies sparked by prominent Freedom of Information requests for research data have had a detrimental effect on institutional reputations and brought the risks of poor data management into sharp focus. Concurrently 'data driven' technologies have reshaped the research process and demonstrated benefits of scale and impact in a growing number of disciplines.

Reflecting the broader changes noted above, the JISC-funded Digital Curation Centre (DCC) supports the UK higher education community to manage, curate and preserve digital material. Most recently, DCC effort has been focused on managing research data. We distinguish RDM from preservation by the former's emphasis on verifiable and replicable processes to support research data use from its planning, through its creation and active use, to its point of handover to a repository or archive. These include preservation actions to ensure fitness for access, use and reuse, as described for example in the DCC Curation Lifecycle Model [1]

Research data management represents new demands for HEIs in terms of technical and organisational infrastructure, the provision of specialist data curation skills and long term planning for sustainable services. We are currently working with twenty-one HEIs through our institutional engagement programme to increase their RDM capability in these areas whilst developing a support model that can be redeployed with other UK universities charged with facing what are commonly seen as additional technological and policy challenges.

There are two key outputs from the DCC institutional engagement programme: 1) a model for supporting HEIs to develop their RDM capabilities, i.e. their ability to articulate and achieve RDM objectives; and 2) a transferable RDM toolkit. The support model is outlined in section 3. It involves applying tools to help initiate processes of change in each institution, diagnosing current practice, and implementing redesigned services. The RDM toolkit describes potential HEI services, examples of which are given in section 4. These include exemplars of DMP Online, an online data management planning tool customised for HEIs by using 'institutional templates'. Each of the HEIs the DCC is supporting

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. iPRESS2012, Oct 1-5, 2011, Toronto, ON, Canada. Copyright 2012, Digital Curation Institute, iSchool, University of Toronto.

has agreed to share their experience and to allow others to reuse outputs from our engagement with them.

2. DATA POLICY BACKGROUND

An increasing number of HEIs are developing policies and implementation plans for research data management. These are often guided by funder requirements and codes of good research practice. The Engineering and Physical Sciences Research Council's (EPSRC) policy framework for research data, which was released in May 2011, places the onus on institutions to address research data management. It sets out clear timescales for implementation: research organisations should develop a roadmap to align their policies and processes with EPSRC expectations by 1st May 2012, and be fully compliant with them by 1st May 2015 [2].

Most research councils have released similar policies promoting the effective management and open sharing of research data. The RCUK Common Principles on Data Policy [3] highlight the importance of policies and plans – both at institutional and project-specific level. Importantly, they also confirm that it is appropriate to use public funds to support management and sharing of publicly-funded research data, enabling the development of support infrastructure.

A trend for institutional research data management policies is evident. A number of policies emerged in 2011-2012 and many more are in draft form awaiting approval, as listed by the DCC [4]. These policies frame the institutional governance needed to develop associated infrastructure and embed good practice. The policies tend to be accompanied by guidelines for implementation or more detailed local policies and processes. Data Management Plans (DMPs) written for specific projects or as group guidelines play an important role in this framework. Six of the seven UK research councils expect researchers to submit DMPs in grant proposals, while the seventh (EPSRC) advocates the importance of plans but does not require their submission.

3. A MODEL FOR SUPPORTING HEIs

Our model for supporting HEIs is being refined by implementing it through the engagement process. We first outline the scope of the two main tools we are applying: DAF (Data Asset Framework) and CARDIO (Collaborative Assessment of Research Data Infrastructure and Objectives), which both originate from digital preservation research and development projects. We then describe three business process change stages that we aim to contribute to in each HEI: initiating change; diagnosing data practices; and (re)designing services. We identify the role of DAF, CARDIO and other tools relevant to each stage.

We describe how the engagements fit within the support model, as shown in Figure 1. This comprises two other ongoing activities; evaluation of each engagement, and comparison across them. These result in forthcoming outputs; firstly reports describing and evaluating each engagement, and our comparisons of these across institutions. The latter will document our refined model, based on improved understanding of how best to deploy the DAF and CARDIO tools to develop institutional capabilities, and factors enabling and inhibiting this. The second main output planned is a transferable RDM 'toolkit' of service descriptions, exemplars and good practice guidance that other institutions can deploy. This includes exemplars of support for Data Management Planning, where localised services have been developed.

3.1 Tools for Engagement

Each 'institutional engagement' aims to build the institution's capability by working with them to articulate the need for change, and scope requirements for redesigned services. We envisage institutional services will combine technology with 'soft' infrastructure including training, guidelines, and policies to support these [5], i.e. the changes needed may be at least as much of an organisational nature as a technical one.

We deploy a range of tools and approaches developed through recent collaborative projects. Two DCC tools have supported the initial work: -

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

The *Data Asset Framework (DAF)* 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 gaps. DAF has been piloted in a number of contexts through case studies [7].

3.2 Developing Institutional Infrastructure

Our assumption is that formally structuring and coordinating data management can benefit research. Nevertheless we take the introduction of effective RDM as a rubric for bringing change to a range of highly diverse activities. Sociotechnical research demonstrates the complexity of developing infrastructure in the context of diverse and changing requirements, and the necessity for both short and long-term views to be included in planning this development [e.g. 8]. We see RDM infrastructure development as a process of change that requires input from at least three perspectives; research practice, management, and information systems development. These perspectives may come from an institutions' Library, IT and Research Support functions, as well as from researchers themselves.

Institutional RDM service development can be viewed as an iterative cycle similar to business process redesign. Ideally long-term planning should be encompassed in a process of learning and continuing improvement. Our initial focus is on early stages of process redesign, which we adapt from Kettinger et al's framework [9]; initiating change, diagnosing data practices, and redesigning services[1].

[1] Our current process emphasizes steps two to three in the six stages identified by Kettinger et al [9]; envision, initiate, diagnose, redesign, reconstruct and evaluate.

We characterize the three stages below, indicating relevant DCC methods and tools alongside other examples. The support described is aimed at scoping the redesign of support services and roles. As RDM matures as an element of UK institutions' service provision we anticipate that support for implementing new systems and evaluating services may require further modeling tools to relate these to, for example, enterprise IT architecture (e.g. [10]).

3.2.1 Initiating change

This stage is led by a champion authorised by senior management to form a steering group to scope a project and enlist academic engagement. In several cases, a member of the senior management group responsible for research chairs this, e.g. the Deputy/Pro Vice Chancellor for Research. A steering group would typically consider research strategies, service priorities and technology opportunities, and identify stakeholders, issues and domains to investigate. Having planned and secured the necessary resources for a project, including the human resource to plan and implement change, their initial work is likely to focus on raising stakeholder awareness and obtaining "buy-in".

Engaging senior researchers will be vital given the differences between research and an institution's administrative processes, research practices being more diverse and fluid. Senior management support is also needed to ensure that strategy is aligned with feasible action, given the competing demands for resources and a constrained funding environment. Policy development may be needed to communicate institutional priorities and define responsibilities. Benchmarking to identify capability gaps and analysis of risks and benefits may help the case for change and identify the main goals and success factors.

Methods/tools: DAF, CARDIO, KRDS/I2S2 Benefits Analysis Tool [11].

3.2.2 Diagnosing data practices

The next stage involves profiling data management and sharing norms, roles and values, aiming to identify the main issues encountered by researchers and other service users or stakeholders. Typically a project manager or steering group member with operational responsibility will undertake this work in a series of short studies, involving selected research groups and providers of any relevant existing services such as backup storage or library support.

The aim here is to appreciate enough about current RDM practices, their shaping by disciplinary factors, and usage of available sources of support, to identify the appetite for change, how needs are framed, and the likely barriers to aligning them with strategy and regulatory requirements. The diagnostic stage may therefore include assessment of the awareness of relevant policies, and chart the lifecycle of typical data assets and associated research objects (software, protocols, logs, etc).

Methods/tools: The DAF approach aims to support this form of enquiry into typical data lifeycles, stakeholders involved, and their concerns and priorities. CARDIO complements this by identifying service providers' assessment of current provision. Other tools and methods relevant here include Data Curation Profiles [12], and Stakeholder Profiles [13]. Benefits frameworks may help identify priorities, e.g. the KRDS/I2S2 Benefits Analysis Tool (ibid). Where there is substantial existing support for data archiving and a need for more detailed analysis of workflows, Research Activity Information Development (RAID) diagrams provide a modeling tool to support this [14].

3.2.3 Redesigning research data services

This stage involves the project manager and any operational group working with stakeholders to describe new service options, and their feasibility and desirability. The tools relevant here will

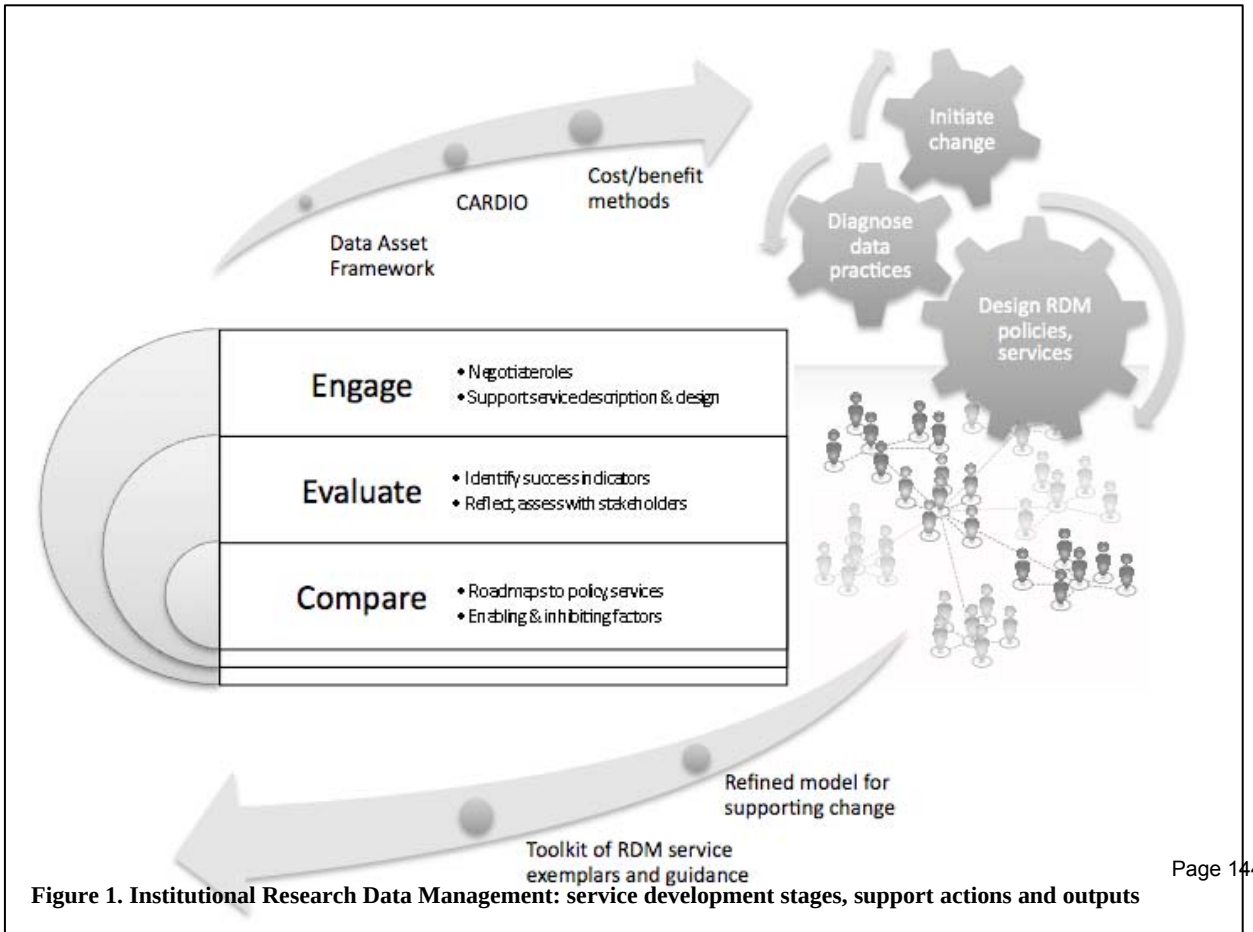

Figure 1. Institutional Research Data Management: service development stages, support actions and outputs

vary with the kind of service proposed. Frameworks for Institution-wide policies and guidelines may take the form of exemplars drawn from other institutions. This is likely to involve senior academics, along with Research Office and or Library (e.g academic liaison) colleagues, in assessing options according to needs scoped using CARDIO or DAF. Similar stakeholders will be involved in defining training needs, and here too the CARDIO and DAF tools should have highlighted the policy areas and RDM concepts that training needs to raise awareness of.

Further tools will be needed when the options for change involve developing new information systems, or include requirements to interoperate with existing systems e.g. institutional repositories, or research information management systems. This will involve business analysts from Library or IT systems areas. Tool support may be needed to articulate new process concepts. This may for example use scenarios to present narrative 'user stories' and use cases. Workflows diagrams (e.g. the RAID method outlined above) and prototypes may help the intended users and stakeholders to compare 'as-is' and 'to-be' processes, whether on cost/benefit or other criteria e.g. research ethics or strategic objectives.

Support for researchers to develop a data management plan (DMP) when applying for research funding is likely to be one component of service provision. DCC provides a tool, *DMP Online* [15] that contributes here by providing templates and guidance to encourage good practice. *DMP Online* originates from a checklist to help researchers meet funders' grant application requirements [16]. The tool can be adapted to individual institutions and our experiences in doing so through the engagement programme provides further lessons in the drivers and barriers to implementing RDM services.

Methods/ tools: RAID diagrams, DMP Online tool, Stakeholder profiles, Soft Systems Methodology [17].

3.3 Evaluating and Comparing Engagements

Each institution provides a mini-case study of factors decisive in shaping institutional research data management; and each offers opportunities to refine the DCC tools and the use cases for delivering these either as generic web-based applications, or as bespoke offerings used with substantial DCC mediation.

Action research methods emphasise learning methodically from involvement in problem solving, and are appropriate given that the DCC programme is funded as capacity building rather than research *per se*. Soft Systems Methodology (SSM) is an action research method consistent with our assumptions about the need for 'soft' infrastructure. While we do not claim to follow the approach rigorously, the important aspects for our purpose are:

- Identifying and engaging with stakeholders who are articulating the need for change
- Appreciating how they frame relevant issues and contexts
- Providing opportunities to articulate feasible and desirable service improvements

To these ends, the authors and colleagues participate in institutions' steering groups, hold workshops with stakeholders to discuss findings, and provide training in good data management practice. To support cross-institutional comparisons the programme holds internal workshops to reflect critically on factors enabling and inhibiting success, across the institutions and our interventions to support them. This also benefits from participation in external workshops held by the JISC Managing Research Data programme, which is funding institutions to conduct similar organisational change and service development projects [18].

Key questions guide our evaluations, whose overall aim is to refine the support model with our stakeholders input, and compare individual engagements. Our key questions are:

1. What stakeholders become engaged in RDM service development, and what new roles are adopted?
2. What are common priorities for RDM services, and enablers and barriers to developing these?
3. How much intermediation is needed to use DAF and CARDIO and how may these best be used in combination?
4. What are our client's and stakeholders' success indicators, and how do they assess our contribution?

The rest of the paper addresses the first two questions, and we conclude on the scope of the RDM service toolkit.

4. PROGRAMME PARTICIPANTS

The engagement programme was promoted to institutions via the DCC's data management roadshows [19]. These are regional events whose main aim is to bring stakeholders together to address institutional RDM issues. The roadshows have encouraged interest, and most of the engagements were initiated through them. For example in a recent roadshow a local institution's, Head of Internet Services, Library Academic Services Manager, and Head of Research Development came together to develop a strategy. They subsequently approached the DCC for assistance and we are defining a programme of support.

The level of interest in the programme has allowed us to establish a balanced portfolio. The twenty-one HEIs currently taking part are spread geographically across the UK and represent a range of university types. Three participants are ancient universities, formed in the 15th and 16th centuries. Another six participants are civic institutions with origins dating from the 19th and early 20th centuries. Eight were formed in the 1960s, while the remainder are former polytechnics that became universities post 1992.

We have sought participation of universities with a variety of research portfolios and strategies. Six of our participants are members of the Russell Group, which "represents the 20 major research-intensive universities of the UK" [20]. Several others are known for particular research strengths and bring these to the portfolio. The more modern institutions focus primarily on teaching but have ambitions to develop their research profile.

The EPSRC policy has been a key impetus for institutions to form working groups with the intention of developing RDM strategies. As we expected these involve a range of services, typically the library, IT and research office. The lead partner in the majority of our engagements is the library. Indeed every engagement has some representation from the library; in cases where they are not leading, library-based staff often undertake the majority of the work. The research office is leading in seven of the cases and is

involved in most of the others. Institutional IT Services are only leading in two of the engagements. Furthermore, IT involvement is lacking in a few other engagements, raising questions about how effectively technical change can be embedded.

5. INITIAL FINDINGS

If the institutions the DCC is supporting can be seen as representative, then UK universities are in the early stages of addressing research data management. Most are scoping requirements and benchmarking current practice to plan future work. We are aware of few institutions where components of an RDM infrastructure or support services are already in place. Most are early in the process of developing services.

The following sections highlight some key areas of activity.

5.1 Research Data Management Policies

Many of the participating institutions have responded to the trend to develop research data management policy. Requests for support have ranged from feedback on drafts to developing policies on their behalf. The DCC has provided a policy briefing [21] in support of this activity, which outlines requirements and summarises different approaches that universities have taken. A number of institutions have looked to the University of Edinburgh's seminal policy developed in 2010-11 [22] and used that as a base from which to adapt.

The DCC capitalized on broader interest in this area by inviting participants to join a JISC Managing Research Data (MRD) policy workshop was held in March 2012, which provided an opportunity to share practice and learn from others. Key discussion points were the degree of specificity needed and the optimum timing of an RDM policy. Questions were raised about the level of detail required of an institutional policy leading to suggestions for more detailed implementation guides and tailored departmental policies. Fears were also expressed about approving RDM policies before the associated infrastructure was in place to make compliance problematic.

Only three participants have RDM policies that pre-date DCC involvement. In these cases the emphasis of our work is on policy implementation. Pilot studies are being run with researchers at one institution to see how easily they can write a data management plan and deposit data for preservation and sharing, as outlined in the policy.

5.1.1 Example A: Policy development

One participating university created a Research Services librarian post in 2011. This post aimed to support researchers, in line with the institution's ambition to be a leading modern university for research. The person recruited was tasked to lead the University's RDM initiatives. As in other institutions, the EPSRC expectations were a driving force.

The initial task in this university was to develop an RDM policy. Existing policies were reviewed in November 2011. A first draft was largely based on the University of Edinburgh policy, with additions to define further responsibilities and agree periodic review dates for data retention. Feedback from a small focus group was positive; researchers sought clarification on the scope and wanted practical guidance for implementation. There was also a desire that the policy should be supportive rather than strongly enforced. With researcher support, the policy was put forward to the Research & Knowledge Exchange Committee and approved.

The policy development and approval at this institution, a relatively new university, took four months. In part this is due to existing examples that could be repurposed, drastically reducing the effort needed in composition. The process of approval was also far simpler than in older universities, which tend to have various committee levels that need to be passed.

5.2 Roadmaps and Strategy Development

The EPSRC's policy [2] places a number of expectations on institutions. They must ensure awareness of the policy and regulatory framework for RDM, identify internal data holdings, publish metadata about these, and provide infrastructure to preserve them. The policy calls for long-term commitment to preservation; institutions are required to keep selected datasets accessible for at least ten years from the end of any embargo period, or from the date of the last third party access request. Institutions must also define responsibilities for curation activity across the DCC Curation Lifecycle.

The EPSRC expects 'roadmaps' to plan RDM infrastructure and services and ensure compliance with their expectations. This has provided the context for the DCC to help participating institutions scope a response. Institutions should define the content and format of the roadmap and initially self-assess their compliance. However the EPSRC has made clear that future funding may depend on inspection and compliance. This has provided an impetus for our work with RDM steering groups.

This work has drawn primarily on the DAF (Data Assessment Framework) and CARDIO (Collaborative Assessment of Research Data Infrastructure and Objectives) tools. Typically steering groups have preferred CARDIO where their institutions have a range of relevant services in place. Other steering groups have preferred to conduct DAF surveys or interviews to gather evidence of current awareness and needs. These have been carried out through pilot groups, identified with varying degrees of DCC support. The pilots provide evidence for developing roadmaps and policy, and a model for the steering groups to apply with further groups across their institution.

DAF questionnaires have been tailored to suit institutional circumstances; in some cases they have been used online and as the basis for structured interviews; in others as topic guides for semi-structured interviews. In some cases steering group members have undertaken the interviews themselves, with DCC advising on questions and format, and in others they have shadowed the DCC staff doing interviews and, having gained familiarity with the topics and structure, taken a more active role in later interviews.

We have provided workshops at the beginning of these pilot studies, often combining RDM training and awareness raising sessions with introductions to the DAF approach. We also use workshops towards the end to communicate and consolidate results. CARDIO has been used for both purposes; some institutions have opted to use it to benchmark service provision before further investigation, others to take stock of the results of the investigation.

5.2.1 Example B: Roadmap development

One institution's steering group is led by its Research Office and Records Management staff. Our role was to propose a roadmap format and gather evidence for initial self-assessments. Initially this involved helping to define pilot groups in two faculties, and then carrying out DAF interviews with researchers at various levels of seniority, from doctoral students to research group leaders. These profiled current practice, gauged demands for change, and informed a CARDIO gap analysis against EPSRC expectations. The interviews with researchers and support staff across faculties also highlighted gaps between expectations that other funders place directly on researchers themselves, and the support available to help researchers meet these.

It proved useful to organise the roadmap under the headings of training, policy development, service development and policy implementation, and finance. This helped separate tasks that could be accomplished in the short term, from others requiring additional roles and resources. Short-term requirements included basic RDM training to be embedded in postgraduate training. Longer term requirements included systems for cataloguing active research data at faculty and/or research group level, guidelines and processes for appraising and selecting material of long-term value, and identifying the appropriate place to deposit/ preserve it or ensure appropriate disposal.

5.3 Data Management Planning

The DCC web-based tool to assist in this process, DMP Online, has three main functions: to help create and maintain different versions of DMPs; to provide useful guidance on data management issues and how to meet research funders' requirements; and to export useful plans in a variety of formats. The tool draws upon the DCC's analysis of funders' data requirements to help project teams create two iterations of a data management plan: an 'application' stage plan and a 'funded' plan.

Several of the institutions in the programme have asked for a tailored version of DMP Online. This enables universities to add customised guidance, such as links to relevant webpages and contact details for support staff. A new feature in v3.0 of the tool is the ability to provide suggested answers: universities can compose text for inclusion in cases where generic provision is in place, such as central storage and backup. Customised versions of DMP Online incorporate the institution's logo and can be branded to apply relevant design and URLs so they are seen as an institutional service.

5.3.1 Example C: Customising DMP Online

At one participating university DCC support is part of the institution's IT Transformation project, which is addressing various aspects of research data management, including storage and tools. Some preliminary work on data management planning was undertaken by a JISC-funded project in a research centre in the university. This provided the catalyst for a customisation of DMP Online.

A preliminary meeting was held early in 2012 to discuss requirements with the project manager. The process of customization was explained and a schedule agreed. An implementation team at the institution has documented requirements and produced an institutional template based on the elements of the DCC Checklist, which they wish to include together with details of local support. The DCC has input this information to create the template in the tool, and supported ongoing user testing. Training materials are being developed to suit this institution's context and a launch is planned for 2012.

5.4 Managing Research Data Storage

Managing storage is a primary concern for researchers, and as such is high on the list of priorities for universities. Activity is typically focused on providing sufficient quantities of research data storage. Tools to enable data sharing with external collaborators and version control are also sought. Analogies are often made with Dropbox when describing requirements [23].

Significant developments in this area are being made in the wider community. The DataFlow project at the University of Oxford [24] is one of a range of RDM applications resourced by the Higher Education Funding Council for England (HEFCE), as potential cloud-based services for universities. DataFlow is a two-stage data management infrastructure intended to make it easy for researchers to work with, annotate, share, publish, and permanently store their research data. There are two components: DataStage, a secure, local file management system with private shared and collaborative directories, and DataBank, a scalable data repository designed for institutional deployment. Several of the universities DCC is supporting have flagged an interest in piloting DataFlow.

5.4.1 Example D: Data storage strategy

At one institution a Vice Principal convened two working groups to progress their RDM initiatives: one on research data management and one on research data storage. The research data storage working group identified requirements for a cross-platform file store, accessibility for external collaborators, and provision for backup and synchronisation. Requirements were also identified for services to deliver data archiving and federated data storage.

A business case was made and resources released to purchase infrastructure and develop support services. The DCC has assisted the working group to develop a list of existing and proposed services. Pilot studies are planned to test the different ways forward. The expectation is that existing provision will be extended to allocate a nominal 0.5TB per researcher, with provision co-ordinated at local level.

5.5 Guidance and Training

DAF and CARDIO studies have uncovered a discrepancy between existing support provision and awareness of this. In many cases collating details of existing services and improving their presentation presents a 'quick win'. This was done on the JISC-funded Incremental project at the Universities of Cambridge and Glasgow, and provides a useful model for redeployment [25]. Short, simple guidance tends to be called for, as data management can seem overwhelming if presented in a technical way.

Training of some kind features in over 25% of the engagements. There are two key areas of interest: disciplinary courses for PhD students and professional training to re-skill research support staff. Our emphasis is on extensively reusing existing resources. The DCC's DC101 course [26] and Data Intelligence 4 Librarians

[27] by the 3TU consortium in the Netherlands are both targeted at research support staff. The JISC RDMTrain projects produced disciplinary courses [28] and the UK Data Archive has also produced training materials for researchers [29].

The DCC provides bespoke training courses by adapting relevant resources to specific institutional needs. Requirements currently being addressed include provision for one institution's academic liaison librarians to introduce RDM to researchers; and in another institution providing content for PhD training in Health and Life Sciences.

5.5.1 Example E: Training development

One participating university has also been running a JISC MRD infrastructure project. In collaboration with that team, we have supported a number of training initiatives. Training is run via the Doctoral Training Centre, with the hope that by catching young researchers early, you can instill good data management habits before they start to make bad ones. We trained the most recent cohort at the beginning of the academic year and they have supported one another since. The training gave a grounding in research data management and used data management plans as a vehicle to put the principles into practice. The PhD students trialed a number of DMP templates to see which was most appropriate to develop a plan to guide their work.

6. CONCLUSIONS AND NEXT STEPS

The DCC model for supporting institutions to build RDM capacity is working well. In all cases a plan of action has been developed with a steering group and is in the process of being delivered. For the majority of institutions this has involved diagnosing current practice to define requirements, as most were unaware of their current position at the outset. A few institutions are nearing completion of the DCC engagement, having indicated they feel equipped to continue development themselves.

In many cases, the initial stages have taken some time to build momentum, as the process of reaching consensus and initiating change can be daunting. However, progress has been far quicker in some institutions than others. This appears to be due to a range of factors. A few institutions have committed resource to research data management and funded a position to spearhead activity and build momentum. There also seems to be a quicker process of change in smaller, more modern institutions. This could in part be due to their structure: fewer levels of hierarchy make it easier to raise ideas and elicit approval. Cultural factors may also be at play: smaller scales can make it easier to engage the research community and there appears to be a greater willingness amongst researchers to work with central services.

Some approaches have worked particularly well, such as our focus on engaging early career researchers in training, in the expectation that they will filter change upwards as they permanently adopt good data practice as part of their routine research process. With policies defined and the benefits explained, institutions are also beginning to grapple with the creation of business plans designed to ensure that the necessary technical and human infrastructures are sustainable. We have found the principal concerns across participating institutions' steering groups to be similar. The main indicators of success for them are the formulation of roadmaps to address compliance requirements, which are common to all UK universities (e.g. [30]), and 'quick wins' in terms of responses to researchers' demand for clear guidance and easily managed storage provision.

Despite the DCC emphasis on providing generic web-based solutions, engagement demands flexibility and adaptation to local contexts. Most of the usage of DAF and CARDIO has been with our mediation, and this has enabled us to identify needs to improve the flexibility and integration of online tools to support this. Evaluation to prioritise specific improvements in tools and methods (e.g. workshop formats) is ongoing, comprising telephone interviews with participating stakeholders, and usability assessments of the online elements of support provision.

Our next steps include cross-institutional surveys on the needs for support in policy compliance, and the degree to which involvement in our programme has supported this. A number of important differences have already emerged through cross-site comparison. Requirements for support vary: some researchers create vast quantities of complex data and require improved storage management to make analysis scalable. For others, the challenges are more in the heterogeneity of data form. Attitudes to data sharing set others apart: those working with human subjects require tightly controlled access, whereas other groups have adopted a culture of data sharing and demand easier external collaboration. Requirements can be diverse across and within disciplines, so a flexible approach is needed.

From a data curation perspective one should not exaggerate the differences. Despite them we find that similar issues apply in supporting data management: policy development and planning, training and guidance, data management planning, managing storage for active research data, data evaluation/appraisal, gathering and publishing metadata, identifying relevant external repositories, choosing repository platforms, systems integration, managing data access and citation, and making the case for long-term sustainability. Many of these issues overlap with preservation, and in supporting active research data management we continue to draw lessons from the preservation community.

On a national level, these are still relatively early days in the change process. Continued support will be needed over the coming years as pilot projects transition into embedded services. For the DCC, the formal conclusion of each sixty-day engagement is not the end of our collaboration. Continuity in support is vital to a community that is fluid by nature and notorious for the speed with which initiatives decay when the driving force is removed before the achievement of critical mass.

The outputs of the DCC engagement programme are adding to a growing body of exemplars that can be repurposed. Parallel work in the JISC MRD programmes, data centres and RDM initiatives in a number of UK universities are similarly providing RDM service exemplars and outputs that can be repurposed. The key for institutions is to draw relevant aspects from these examples, which suit their research culture and environment. The DCC engagement programme aims to provide an adaptive framework for doing this. We hope to refine and share this framework beyond the borders of the engagement programme as a model for other HEIs to improve their research data management practice.

7. ACKNOWLEDGMENTS

We acknowledge the support of the Higher Education Funding Council for England (HEFCE), and the JISC via the Universities Modernisation Fund. Our thanks are also extended to the various

universities that have participated. Final reports from the programme will name the institutions concerned and incorporate their comments. We also thank three anonymous reviewers.

This paper represents work undertaken by a DCC team comprising Brian Aitken, Alex Ball, Michael Day, Martin Donnelly, Monica Duke, Marieke Guy, Patrick McCann, Andrew McHugh, Kerry Miller, Jonathan Rans, and the authors.

8. REFERENCES

[1] Higgins, S. 2008. The DCC Curation Lifecycle Model. International Journal of Digital Curation 3 (1)

[2] EPSRC. 2011. Policy Framework on Research Data: Expectations. http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx

[3] Research Councils UK. 2011. Common Principles on Data Policy. http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

[4] Digital Curation Centre. 2012. Institutional Data Policies. http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies

[5] Ward, C., Freiman, L. Jones, S. Molloy, L. and Snow, K. 2011. Making Sense: Talking Data Management with Researchers. International Journal of Digital Curation 6 (2) (October 7). doi:10.2218/ijdc.v6i2.202. http://www.ijdc.net/index.php/ijdc/article/view/197

[6] Kenney, A. and McGovern, N. 2005. The Three-Legged Stool: Institutional Response to Digital Preservation. II Convocatoria del Coloquio de marzo. Cuba. http://www.library.cornell.edu/iris/dpo/docs/Cuba-ark-nym_final.ppt

[7] Jones, S., A. Ball, and Ç Ekmekcioglu. 2008. The Data Audit Framework: a First Step in the Data Management Challenge. International Journal of Digital Curation 3 (2): 112–120.

[8] Whyte, A. 2012. Emerging infrastructure and services for research data management and curation in the UK and Europe. In Managing Research Data, G. Pryor, Ed. Facet Publishing, London. 173-204.

[9] Kettinger, W. J., Teng, J.T.C and Guha, S. 1997. Business Process Change: A Study of Methodologies, Techniques, and Tools. MIS Quarterly 21 (1) (March 1): 55–80. doi:10.2307/249742.

[10] Becker, C., Antunes, G., Barateiro, J. and Vieira, R. A Capability Model for Digital Preservation. In Proc. iPRES 2011, 2011.

[11] Beagrie, N. 2011. KRDS/I2S2 Benefits Analysis Tool. http://beagrie.com/krds-i2s2.php

[12] Witt, M., Carlson, J. Brandt, D. S and Cragin. M. H 2009. Constructing Data Curation Profiles. International Journal of Digital Curation 4 (3).

[13] Michener, W. K., Allard, S. Budden, A. Cook, R.B, Douglass, K. Frame, M. Kelling, S., Koskela, R., Tenopir, C., and David A. Vieglais, D.A.. Participatory Design of DataONE—Enabling Cyberinfrastructure for the Biological and Environmental Sciences. Ecological Informatics (0). doi:10.1016/j.ecoinf.2011.08.007. http://www.sciencedirect.com/science/article/pii/S1574954111000768

[14] Darlington, M., Ball, A. Howard, T, Culley, S. and McMahon. C. 2011. RAID Associative Tool Requirements Specification (February 23). http://opus.bath.ac.uk/22811/

[15] Donnelly, M., Jones, S and Pattenden-Fail, J.W. 2010. DMP Online: The Digital Curation Centre's Web-based Tool for Creating, Maintaining and Exporting Data Management Plans. International Journal of Digital Curation 5 (1) (June 22): 187–193. doi:10.2218/ijdc.v5i1.152.

[16] Digital Curation Centre. 2011. Checklist for a Data Management Plan http://www.dcc.ac.uk/webfm_send/431

[17] Checkland, P., and Poulter. J. 2010. Soft Systems Methodology. In Systems Approaches to Managing Change: A Practical Guide, Ed. Reynolds, M. and Holwell, S. London: Springer London. 191–242. http://www.springerlink.com.ezproxy.webfeat.lib.ed.ac.uk/content/p35875774m64g2l2/.

[18] Joint Information Systems Council. 2011. Managing Research Data Programme 2011-13. http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

[19] Digital Curation Centre. 2011. Data Management Roadshows http://www.dcc.ac.uk/events/data-management-roadshows

[20] Russell Group. 2012. About Us. http://www.russellgroup.ac.uk/about-russell-group/

[21] Digital Curation Centre. 2011. Research data policy briefing http://www.dcc.ac.uk/resources/policy-and-legal

[22] University of Edinburgh. 2011. Research Data Management Policy. http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy

[23] Cope, Jez. 2012. MRD Hack Days: File backup, sync and versioning, or "The Academic Dropbox" http://blogs.bath.ac.uk/research360/2012/05/mrd-hack-days-file-backup-sync-and-versioning-or-the-academic-dropbox

[24] DataFlow. 2012. What we're doing http://www.dataflow.ox.ac.uk/index.php/datastage/users/researchers

[25] University of Cambridge. 2010. Incremental Scoping Study Report and Implementation Plan. http://www.lib.cam.ac.uk/preservation/incremental/news.html

[26] Digital Curation 101 http://www.dcc.ac.uk/training/dc-101

[27] Data Intelligence 4 Librarians http://dataintelligence.3tu.nl/en/home

[28] JISC RDMTrain outputs http://www.dcc.ac.uk/training/train-trainer/disciplinary-rdm-training/disciplinary-rdm-training

[29] Managing and Sharing data course http://www.data-archive.ac.uk/create-manage/training-resources

[30] Research Councils UK. 2009. RCUK Policy and Code of Conduct on the Governance of Good Research Conduct. http://www.rcuk.ac.uk/review/grc/default.htm