The DCC's Institutional Engagements: Raising Research Data Management Capacity in UK Higher Education

by Martin Donnelly

Research Data Access & Preservation

EDITOR'S SUMMARY

The United Kingdom's Digital Curation Centre (DCC), established in 2004, is the country's digital preservation and data management authority. From 2011 to mid-2013 the DCC's Institutional Engagement program provided support for digital preservation and data management to 20 universities. The initiative was stimulated by the realization that research data leads to broader advances when shared, as well as by pressure from funders to implement formal data management practices. Experience with the first cohort of universities revealed differences in readiness for data management, the importance of a local champion and support staff and the need for dedicated funding. Unfamiliar but necessary working relationships emerged, as did the need to align institution requirements with funder priorities. Common challenges were planning for effective data management, data storage and training for researchers and staff. The DCC will apply the lessons learned with the first cohort to other higher educational institutions, adjusting its approach to serve each university's specific needs while promoting the shared purpose.

KEYWORDS

data curation

research data sets

digital object preservation

colleges and universities

United Kingdom

training

Martin Donnelly is the senior institutional support officer at the Digital Curation Centre, based at the University of Edinburgh, with particular interests in data management planning, policy and arts data. His academic specialty is in digital humanities/cultural heritage computing, and he previously worked at Edinburgh College of Art and at the University of Glasgow. He can be reached at martin.donnelly<at>ed.ac.uk.

The Digital Curation Centre (DCC) is a United Kingdom-supported service and center of excellence for digital preservation and data management founded in 2004. This article gives an overview of the DCC's Higher Education Funding Council for England (HEFCE) funded institutional engagement program (2011-2013), wherein we worked intensively with over 20 UK higher education institutions (HEIs) to identify, describe and raise levels of awareness and capability in institutional research data management. The set of engagements involved HEIs from all constituent countries of the United Kingdom and covered a range of institutional types, from ancient, research-intensive universities to newer universities seeking to raise their research profiles to specialist institutions with a focus in particular subject areas such as economics or art.

The Institutional Engagement Program: An Overview

Between Spring 2011 and Summer 2013, the DCC undertook a significant outreach program designed to assist a cohort of individual universities in the development of their research data management capabilities. Funded by the HEFCE's Universities Modernization Fund, in sympathy with the DCC's core funding from Jisc, this became known as the Institutional Engagement (IE) Programme.

This work shared as its bedrock the view subsequently expressed in the Royal Society's *Science as an Open Enterprise* report ([1], p.8) that "a shift away from a research culture in which data is viewed as a private preserve" is essential to achieving improvements in the exploitation of research. Such cultural change requires the involvement and collaboration of numerous stakeholder groups and the joining together of what could be thought of as work silos. The work does not stop at the gates of the university, though: research funders, charities, government agencies and learned societies are all also involved.

CONTENTS

Context

A number of factors influenced this work, including the increase in data-related requirements and expectations of research funders, publishers and the government. At the highest level, the UK government, via a white paper published in December 2011, has exerted increased pressure on their seven public funding councils to implement and enforce the data management policies and expectations that are currently in place.

Over and above this white paper, the Engineering and Physical Sciences Research Council (EPSRC) – the largest of the research councils in terms of monies disbursed – has informed all research institutions that they are each expected to prepare a data management roadmap by May 2012 and to have a functioning research data infrastructure in place by May 2015, galvanizing many senior managers, researchers and research support staff to take prompt action.

Other funders, notably the Medical Research Council and Economic and Social Research Council, have tightened up their existing regulations, and carried out quality checking exercises on the data management plans received at the application stage [2]. All of these developments point at a trend towards increasing regulation and requirements and expectations of higher quality data management.

The First IE Program

Cohort. At the very beginning of the program, we were able to help the limited number of universities that were interested in participating and met the basic criteria for the program, but – as word got around – demand to participate increased, and by the mid-point we had a waiting list of institutions keen to work with us. These institutions are now being contacted with a view to taking part in a follow-up program beginning in Summer 2013.

The first cohort comprised a variety of institutions, ranging from ancient, research-intensive universities such as Glasgow and Edinburgh to newer (post-1992) universities with smaller research portfolios to specialist institutions such as the London School of Economics and the University of the Arts London. We sought to work with at least one institution from each of the four countries that make up the United Kingdom: England, Scotland, Wales and Northern Ireland.

We initially calculated that the available funding would enable 18 full engagements, with an allocation of 60 days of DCC effort provided free for each, drawn from a mixture of staff based at each of the three DCC sites: Edinburgh, Bath and Glasgow. In practice, however, given the pace of the individual engagements, which in some cases were becoming extended over a lengthy period, we were able to extend this assistance to a few more institutions based on their particular circumstances.

Process. Each engagement was tailored to meet the specific needs of the institution in question, with the DCC providing in-depth and concentrated support to a community of institutions already known to be underprepared, in general, to face the growing challenge of research data management.

Following initial contact and agreement from a senior champion, the DCC and the institution agreed on the key contact people to manage the process from both sides. In most cases, task forces or working groups were assembled, and in some instances a cycle of business process redesign was put in motion. Three key perspectives were encompassed: that of the research practitioners themselves; support staff from research offices, libraries and IT departments; and the senior managers with their hands on the purse strings. (This situation was reminiscent of Cornell's "three-legged stool" model, wherein technology, resources and organization need to be roughly equal in order to keep the chair upright [3].)

It quickly became apparent that the engagements needed authority, clarity and drive if they were to be successful. Without the ongoing backing of a senior figure, research support staff struggled to motivate or engage researchers. Without dedicated resourcing, this work often took a backseat to other pressing concerns, not least of which was the Research Excellence

FIGURE 1. Developing institutional infrastructure.
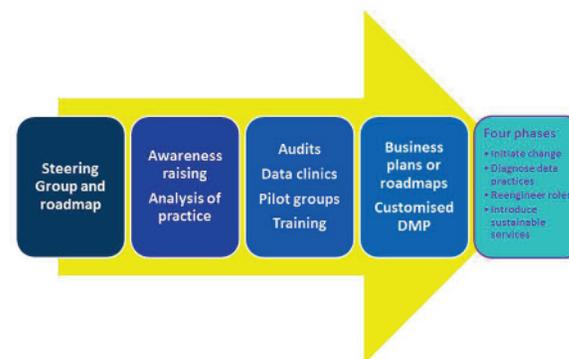
Framework exercise, the semi-regular mechanism by which the UK government determines how much funding to provide to each university.

This multi-stakeholder endeavor led to the creation of often unfamiliar collaborations and assemblages, as central services became responsible for issues that had previously been completely left to the researchers to handle, and, similarly, researchers became

FIGURE 2. IE teams

obliged to work more closely with administrative departments that they had been more accustomed to keeping at arm's length.

Activities Undertaken. When developing work plans for each engagement, we sought to focus on desirability, achievability and sustainability. We found that early-stage funder analysis was useful, determining which funders were most important to the institution in question – both now and in the strategic future – and identifying the requirements that the institution would have to meet in order to safeguard this income stream.

The approaches taken to developing these new, data-centric competencies covered a range of contexts, aims, methodologies and processes. Work involved in the engagements ranged from developing research data management roadmaps and policies to identifying training and support needs to trialling and customizing tools such as DMPonline, the Data Asset Framework (DAF) and CARDIO to integrate within the existing technical infrastructures.

Each of the work programs was different, but we found that three issues predominated: planning, storage and training. Six of the seven state-owned Research Councils-UK (RCUK) funders require data management plans (or an equivalent) at the application stage, and one, the Natural Environment
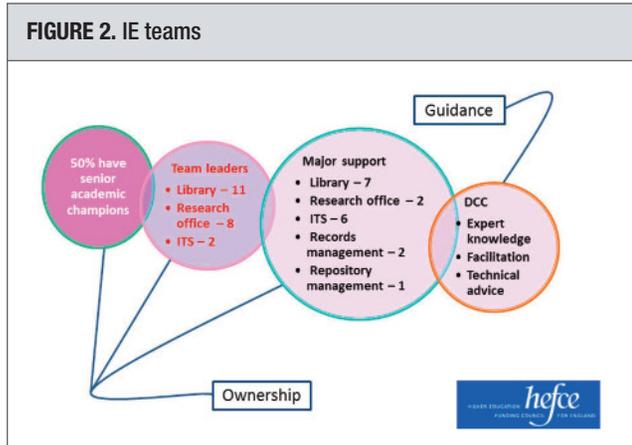
Research Council (NERC), requires two versions: a minimal plan at application and a jointly-agreed upon plan once funding is granted. The DCC's DMPonline data management planning tool was therefore popular with the institutions, enabling the creation of tailored and branded versions that incorporated boilerplate responses, if desired, with institutional plan templates, as well as helpful links to local web pages, training materials, support contacts.

On the storage side, universities frequently came up against the problem of providing sufficient capacity for research data, while also competing with commercial offerings such as Microsoft Skydrive, as well as tools to enable sharing with collaborators such as Dropbox. Universities also struggled to provide advice on the selection of storage options such as onsite repository, national services or cloud services and advice on how to integrate relevant storage-focused initiatives, such as the Jisc-funded DataStage and DataShare projects.

FIGURE 3. IE activities
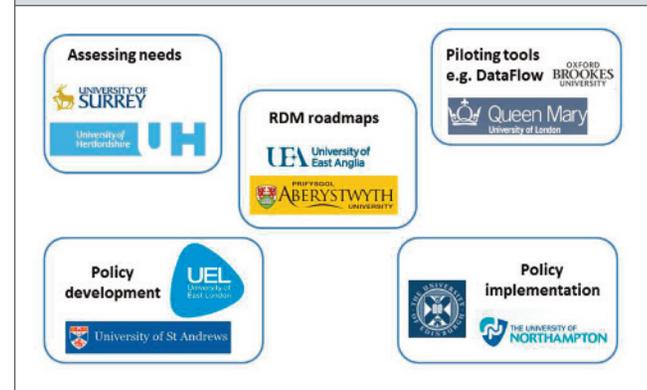
FIGURE 4. Imperatives of guidance and training
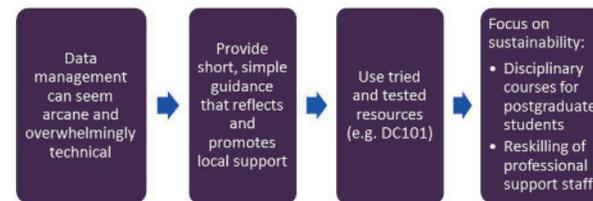
The first of the EPSRC data-related expectations, which underpins any working data infrastructure, is an appropriate level of awareness and accompanying training resources, so that colleagues

know what is expected of them. It is often said that it is neither necessary nor desirable for every researcher to become an expert in all aspects of research data management, but rather they need a decent grounding in basic principles (such as keeping a regular backup of anything significant and making sure it is not held in the same building as the master copy) and to be comfortable in seeking advice or help from the colleagues best placed to provide it. These are largely human infrastructure issues, and frequent communication is necessary to ensure they stay afloat in people's minds.

Lessons Learned. At the time of writing (June 2013) we are finishing off the few engagements from the first cohort and beginning the next batch. The lessons and examples of the first cohort will be shared with the community via a synthesis report that is currently in preparation, and details will be available via www.dcc.ac.uk/community/institutional-engagements soon.

The chief interim observations are:

- Universities are generally in the early stages of scoping research data management (RDM) needs and obtaining benchmarks.
- Few have the effective components for RDM services or infrastructure in place at the moment.
- Management tend to be more concerned with meeting funder expectations than with a desire for sustainable infrastructure.
- There is confusion over the sequence for producing strategies, plans and policies.
- The gulf between early adopters and late entrants is widening.

The Future for These Engagements

With around 140 higher educational institutions in the United Kingdom and only a little over 20 involved in the first cohort, there is, of course, much work still to be done and more demand than the DCC can currently meet. We have, therefore, realigned our program to reflect the widening spectrum of need, offering specific curation techniques to enable infrastructure development, together with sociotechnical support, from advocacy to skills reengineering to organizational repositioning.

DCC staff will continue to be engaged as expert hands-on consultants delivering specific tasks, but the one-size-fits-all approach of 60 days has

been done away with; instead, universities will make requests for specific, defined work packages which will be agreed upon by the DCC and a senior member of the university. Upon satisfactory completion of one piece of work, we can move on to another at the same institution, provided it offers a clear benefit to the institution or the wider research community.

Finally, HEIs will be required to demonstrate commitment to maintain the engagements, and if progress is not being made we have more frequent stop-go points to enable us to provide support where it is most needed and will be most valuable.

Acknowledgements

Resources Mentioned in the Article

[1] Royal Society Science Policy Centre. (2012). Science as an open enterprise. London: The Centre. Retrieved June 13, 2013, from http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.

[2] A special Research Data Management Forum event held in April 2013 brought together a number of representatives of the UK research funders to enable them to go on the record about what aspects of research data management infrastructure (human and technical) they were and were not happy to fund via their grants. See Graham Pryor's blogpost, *A conversation with the funders.* Retrieved June 13, 2013, from www.dcc.ac.uk/blog/conversation-funders.

[3] Kenney, A.R., & McGovern, N.Y. (2005). The three-legged stool: Institutional Response to Digital Preservation (PowerPoint slides). Presented at II Convocatoria del Coloquio de marzo, Cuba. Retrieved June 13, 2013 from www.library.cornell.edu/iris/dpo/docs/Cuba-ark-nym_final.ppt.