The International Journal of Digital Curation
Volume 2, Issue 8 | 2013

Meeting the Data Management Compliance Challenge: Funder Expectations and Institutional Reality

Catherine Pink,
Research Data Scientist,
University of Bath

Abstract

In common with many global research funding agencies, in 2011 the UK Engineering and Physical Sciences Research Council (EPSRC) published its Policy Framework on Research Data along with a mandate that institutions be fully compliant with the policy by May 2015. The University of Bath has a strong applied science and engineering research focus and, as such, the EPSRC is a major funder of the university's research. In this paper, the Jisc-funded Research360 project shares its experience in developing the infrastructure required to enable a research-intensive institution to achieve full compliance with a particular funder's policy, in such a way as to support the varied data management needs of both the University of Bath and its external stakeholders. A key feature of the Research360 project was to ensure that after the project's completion in summer 2013 the newly developed data management infrastructure would be maintained up to and beyond the EPSRC's 2015 deadline. Central to these plans was the 'University of Bath Roadmap for EPSRC', which was identified as an exemplar response by the EPSRC. This paper explores how a roadmap designed to meet a single funder's requirements can be compatible with the strategic goals of an institution. Also discussed is how the project worked with Charles Beagrie Ltd to develop a supporting business case, thus ensuring implementation of these long-term objectives. This paper describes how two new data management roles, the Institutional Data Scientist and Technical Data Coordinator, have contributed to delivery of the Research360 project and the importance of these new types of cross-institutional roles for embedding a new data management infrastructure within an institution. Finally, the experience of developing a new institutional data policy is shared. This policy represents a particular example of the need to reconcile a funder's expectations with the needs of individual researchers and their collaborators.

International Journal of Digital Curation (2013), *2*(8), 157–171. http://dx.doi.org/10.2218/ijdc.v2i8.280

 doi:10.2218/ijdc.v2i8.280

Introduction

The University of Bath is a comparatively small research-intensive UK university with an international reputation as a top-ten university[1]. Links between research and commerce were written into the University of Bath's Charter[2] when it was incorporated, resulting in much collaborative research between the university and industrial, commercial and public sector partners. The University of Bath has a strong applied science and engineering research focus and, as such, the Engineering and Physical Science Research Council (EPSRC) is a major funder of research at the University of Bath.

In 2011 the EPSRC, one of six research councils and other bodies that fund primary research in the UK, published its new Policy Framework on Research Data[3]. This policy included nine expectations[4] covering all aspects of data management including the requirement for institutional policies, data and metadata publication, restrictions on access, length of preservation, persistent identifiers, non-digital data, and resourcing. In a change from the approach taken by many other UK funding bodies, responsibility for compliance was placed on the institution rather than on individual researchers. The EPSRC set two deadlines for the institutions that it funds: by May 2012 institutions were to have a roadmap in place, setting out how they planned to comply with the EPSRC's policy. Full compliance with the expectations is then required by May 2015. What made UK universities take particular note of this policy was the EPSRC's assertion that non-compliance would incur sanctions, which could ultimately include ineligibility for future EPSRC funding.

Due to the importance of EPSRC funding to its research effort, the University of Bath took the EPSRC's new policy framework extremely seriously. The university had already established a Research Data Steering Group to advise on data management issues across the institution. In 2011, this group successfully applied for funding to establish a project, Research360, which would initiate and pilot the work required to achieve full compliance. Research360 was an 18-month project funded by Jisc's 2011-2013 Managing Research Data programme[5], and was structured around meeting each of EPSRC's nine expectations.

In this paper, the Research360 project shares its experience in starting to develop a new data management infrastructure required to enable a research-intensive institution to achieve full compliance with a particular funder's policy, whilst simultaneously supporting the interests of the university and its external collaborators. The paper

[1] The University of Bath was the Sunday Times Higher Education University of the Year for 2012-2013 and is currently ranked first for student satisfaction and third in the 2013 Sunday Times University Guide.

[2] University of Bath Charter of Incorporation: http://www.bath.ac.uk/about/organisation/governance/statutes/

[3] EPSRC Policy Framework on Research Data: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx

[4] EPSRC's nine expectations: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx

[5] Jisc Management Research Data programme: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

focuses on four essential components of this infrastructure: the roadmap required by EPSRC and a business case to support its implementation; two new data roles established to deliver the project; and the creation of a new data management policy.

'Roadmap for EPSRC' Development

The first step towards meeting the EPSRC's expectations was development of a roadmap setting out how full compliance with the policy framework would be achieved. The 'University of Bath Roadmap for EPSRC: Compliance with Research Data Management Expectations' (Lyon and Pink, 2012) was developed by the Research360 project on behalf of the university's Research Data Steering Group. The roadmap was originally based on Monash University's influential 'Research Data Management Strategy and Strategic Plan 2012 – 2015' (Beitz, Dharmawardena and Searle, 2012). Monash University's strategy and strategic plan clearly demonstrated how the benefits of well managed research data were aligned with the university's long-term strategic aims, together with a series of 13 goals and associated initiatives, designed to ensure that the university continued to enjoy the benefits of improved research data management.

Taking Monash University's strategy as a starting point, the Research360 project team first aligned five new data management themes with the University of Bath's corporate plan and strategic aims. These themes covered areas such as international reputation, innovation, business planning, capacity and capability, and infrastructure. The next step was to convert the goals and initiatives in Monash University's strategic plan into objectives and actions relevant to the thematic areas specific to the University of Bath. This process involved extensive rewriting, recognising that Monash University and the University of Bath differed in terms of external drivers, funding environment and extent of data management infrastructure already in place. For example, Monash University already had a well-developed data management infrastructure and, as such, the strategy focused on leadership, with a goal to 'maintain and grow' international recognition of leadership in research data management. In contrast, the University of Bath focused on its relationship with external collaborators, with objectives including the specification of research data management requirements in new contracts with research partners. Importantly this process meant that, first and foremost, the final 'Roadmap for EPSRC' would meet the needs of the institution, not just the funder.

The second stage of the development process involved mapping the new Bath-focused objectives to the nine expectations of EPSRC's Policy Framework on Research Data. As part of this process, the Digital Curation Centre's (DCC) series of blog posts (Jones, 2012; DCC, 2012) in advance of the EPSRC's May 2012 deadline prompted the inclusion of contextual information setting out the rationale for the roadmap's development, and a statement of the current position for each expectation. The latter, relating to the University of Bath's position relative to each expectation, was established based on a Data Asset Framework survey that the university undertook during 2011 (Jones, 2011), from the outputs of an initial test of the DCC's CARDIO tool[6] with subject librarians, and from the experience of members of the Research360 project team.

[6] CARDIO: http://cardio.dcc.ac.uk/

 doi:10.2218/ijdc.v2i8.280

A combination of the objective-mapping exercise and the current position statements enabled gaps to be identified and additional objectives to be created. For example, in order to meet the first expectation that the organisation promote internal awareness of the objectives, an additional objective was added to focus activities, information and guidance on a single RDM website. In many cases, activities were also expanded to ensure full compliance with every EPSRC expectation, such as Objective 2.1, which was expanded to include development of a template data access statement for inclusion in published papers. Ultimately, the 13 goals in Monash University's 'Research Data Management Strategy and Strategic Plan 2012 – 2015' were expanded to 22 objectives in the 'University of Bath Roadmap for EPSRC'.

Using this process to develop the roadmap ensured both that it met a specific funder's policy and aligned with the strategic goals of the institution. For example, Objective 1.1 of the roadmap seeks to 'develop the data management skills and knowledge of Bath researchers' by providing training to researchers, including postgraduate research students in Doctoral Training Centres (DTCs) and graduate schools. This objective would ensure that researchers are aware of the EPSRC's policy, of the external regulatory environment and reasons why data might be withheld, thus ensuring compliance with EPSRC Expectation 1[7]. This objective would also form part of the training and development of postgraduate researchers, thus contributing to the University's Research Strategy[8]. Similarly, Objective 7.1 seeks to 'align digital data storage infrastructure with research and data management requirements'. Delivering this objective would include development of a data repository for long-term retention, archiving and accreditation of research data. This objective would therefore ensure that EPSRC-funded research is securely preserved for ten years from the date of last access, thus meeting EPSRC's seventh expectation[9], but would also contribute to strategic investment in 'high-quality research infrastructure, facilities [and] research support' as part of the University's Research Strategy.

Throughout the development of the roadmap, the support of the Pro-Vice-Chancellor (Research) proved vital. The Pro-Vice-Chancellor provided guidance and was able to anticipate what questions were likely to be raised by committees during the approval process. Importantly, the Pro-Vice-Chancellor also acted as a champion of the roadmap at the Vice-Chancellor's Group, where it was submitted for consideration and final approval. The approval process proved to be a valuable component in the roadmap's development, as it allowed the Vice-Chancellor's Group members to draw upon their extensive experience to provide valuable feedback. For example, their suggestions helped to position the roadmap at a realistic point between minimal compliance with the EPSRC's policy and an

[7] EPSRC Expectation i: "Research organisations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data."

[8] University of Bath Research Strategy: http://www.bath.ac.uk/research/about/strategy/index.html

[9] EPSRC Expectation vii: "Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10-years from the date that any researcher 'privileged access' period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party; all reasonable steps will be take to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK."

ambitious 'gold standard' data management service, which would have been unfeasible to deliver within the time available.

Once approval had been granted, the 'University of Bath Roadmap for EPSRC' was submitted to the EPSRC in time for the May 2012 deadline. Feedback from the EPSRC was positive, with Ben Ryan, EPSCR's Senior Evaluation Manager, commenting that the roadmap was "an excellent example of an appropriate response [that] fully meets our needs for assurance that the University is taking our policy framework on research data seriously."[10]

The Business Case for Data Management

A key aspect of the 'Roadmap for ESPRC' was that responsibility for implementation and management oversight was assigned for each of the 22 objectives. In every case, this responsibility was shared between a number of key stakeholders across the university beyond the initial Research360 project team, ranging from professional service departments and committees to smaller research groups or individuals. This highlighted the extent to which ensuring effective and sustainable management of research data represents a shared responsibility, requiring collaboration between different services within the university. However, in recognition of the demands already placed on these key stakeholders, another outcome of the Research360 project was to develop longer-term strategic plans to ensure that sufficient resource was available to implement the 'Roadmap for EPSRC'. The development of a business case to support investment in research data management was central to these plans. To create this business case, the Research360 project team worked with Charles Beagrie Ltd[11], drawing on its extensive experience in cost/benefit analyses in the areas of data management and digital preservation.

Any business case must demonstrate the benefits that can be derived from investment of additional resource. To articulate the benefits of data management in the context of the collaborative research undertaken at the university, the Research360 project team and Charles Beagrie Ltd built on the Keeping Research Data Safe Benefits Framework[12] to identify benefits both to the university community and to a range of external stakeholders. This was published as 'Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners' (Beagrie and Pink, 2012), which identified how management of research data within the university would benefit both the university community (including researchers, students, professional services) and the institution as well as external partners, (including commercial, public and voluntary sector collaborators, government and society). For example, feedback from industry suggested that it would welcome more open access to research data, as this would provide reference datasets against which new approaches could be tested. Similarly, not-for-profit research partners would benefit from mechanisms that provided enhanced data security and access control in relation to personally sensitive data, as this would encourage more of the public to

[10] News article on the University of Bath Roadmap for EPSRC: http://www.bath.ac.uk/news/2012/06/06/roadmap-epsrc/

[11] Charles Beagrie Ltd: http://www.beagrie.com/

[12] Keeping Research Data Safe: http://www.beagrie.com/krds.php

 doi:10.2218/ijdc.v2i8.280

volunteer as participants in new research. Note that neither of these benefits directly related to any of the EPSRC's expectations.

The business case also demonstrated the importance of data management in the context of a number of key drivers. Central to these drivers was the EPSRC's data policy, so the business case reiterated the university's current level of compliance for each expectation, taken from the 'Roadmap for EPSRC'. In addition, the business case also highlighted the importance of data management in the context of the 2014 Research Excellence Framework (REF2014). To support this, the Research360 project and Charles Beagrie Ltd prepared an overview of how research data and data management might contribute to the three elements of REF2014: research outputs, impact, and environment (Beagrie, McKen and Pink, 2012). These guidelines recognised that 'it is the research activity itself and its impact that is the focus of the REF and the universities' submissions'. However, the university's compliance with the EPSRC expectations would benefit REF submissions and similar exercises, since it would focus on improvements in research data management highlighted by the guidelines as 'a support activity enabling excellent research' (Beagrie, McKen and Pink, 2012).

In order to support the transition from project-based activities to an embedded infrastructure, the business case presented the anticipated medium- to long-term costs of data management, based on a series of case studies prepared by Beagrie, Chruszcz and Lavoie (2008), which illustrated data preservation at a number of UK universities. In these case studies, Beagrie, Chruszcz and Lavoie demonstrated that the costs associated with institutional repositories for research data were an order of magnitude greater than costs associated with archiving e-publications alone. Further, they showed that the staff costs tend to be the major component of preservation costs, particularly during repository startup. Accordingly, the business case made a number of recommendations, which included investment in two permanent posts with responsibility for research data management. These posts were originally created as part of the Research360 project and are described below.

Data Management Roles and Responsibilities

In order to meet the ambitious goals of both the Research360 project and extensive funder requirements, a team of staff from key departments across the institution was brought together. This included the Research Publications Librarian, the Research Information Manager, and representatives from Computing Services, the Vice-Chancellor's Office, UKOLN and a cross-faculty academic research centre: the Centre for Sustainable Chemical Technologies (CSCT). The team was coordinated by a new role, that of the Institutional Data Scientist, who was supported by a Technical Data Coordinator.

Institutional Data Scientist

The Institutional Data Scientist was responsible for the coordination and overall delivery of the Research360 project and, as such, the development of the pilot data management infrastructure across the institution. Based in UKOLN[13], the Institutional

[13] UKOLN: http://www.ukoln.ac.uk/

Data Scientist had a cross-departmental role that facilitated communication and coordination among the different internal stakeholders. This was particularly important, as data management brought together a number of activities that were traditionally seen as distinct services provided by different professional service departments, such as grant applications (Research Support Office), data storage (Computing Services), and archive, publication and open access (Library).

A large component of the Data Scientist's role during the Research360 activity was to build a case for continuing with data management activities once the project finished. This involved working closely with other members of the project team and external consultants to develop the roadmap and business case previously described. To support them, the Data Scientist was responsible for collating evidence of demand for data management infrastructure, such as requests for support, improved management of research active data and re-use of project outputs by other institutions. The latter also demonstrated how investment in data management not only facilitated compliance with funder policies and benefited researchers, but also enhanced the university's national and international reputation, particularly as members of the project team were invited to present expert talks at national and international events.

While most of the project team had other core responsibilities as part of their normal roles within the university, the Data Scientist's sole focus on, and interest in, data management meant that they were able to act as a champion of data management. This not only ensured that the project progressed but also provided a central point of contact for all data management queries, both within the professional support services and, more importantly, for researchers. Here, the background of the Data Scientist as a researcher, both in academia and in industry, proved to be beneficial. Direct experience of the research process enabled the Data Scientist to engage with researchers and assure them that their research needs were regarded as paramount. The Data Scientist was able to demonstrate that the new tools, guidance and technical resources being developed were intended to support the research process and enhance activities already undertaken by researchers, and that compliance with funder policies would be an inevitable consequence of the university's investment in data management, rather than its sole motivator.

Assistance with writing data management plans (DMPs) represents a good example of this support. Many funders require submission of DMPs as part of a grant application. Help provided by the Data Scientist went beyond directing researchers to templates and guidance already available or mandated by funders, to include detailed review and enhancement of draft DMPs. In most cases, the Data Scientist was able to meet with researchers to discuss not only their DMP but more general data management concepts. This direct engagement often enabled researchers to improve the storage of their current data, to discover new metadata standards relevant to their discipline (such as MIBBI[14]), or simply to reflect on their proposed methodologies for forthcoming projects by exploring plans for data capture and processing in more detail. Feedback from researchers regarding this level of support has been extremely positive, providing a foundation for the necessary cultural change needed to ensure compliance with funder requirements.

[14] Minimal Information for Biological and Biomedical Investigations: http://mibbi.sourceforge.net/portal.shtml

 doi:10.2218/ijdc.v2i8.280

Technical Data Coordinator

Supporting the Data Scientist was a second full-time data management role, that of the Technical Data Coordinator. The primary role of the Technical Data Coordinator was to provide general research data technology expertise on a number of project areas, including data repository development, virtual research environments and electronic lab notebooks. An important aspect of the role was to provide specific coordination and communication with technical services, including Bath University Computing Services (BUCS).

The Technical Data Coordinator was seconded to the Research360 project from the cross-faculty Centre for Sustainable Chemical Technologies (CSCT), where they previously provided support for the centre's Doctoral Training Centre (DTC). As such, the Technical Data Coordinator had close links with academics in the Faculties of Science and Engineering, which are recipients of the majority of the University's EPSRC funding. In addition, the Technical Data Coordinator was able to use the centre's DTC as a test bed for many of the outputs of the project:- Postgraduate research students from CSCT attended the first pilot data management training workshop, and also trialled a range of data management planning tools.[15] The established relationship between the Technical Data Coordinator and these doctoral students meant that they were willing to provide constructive feedback on draft deliverables, something which contributed substantially to the improvement of these resources for use by other researchers.

One of the primary methods by which compliance with the EPSRC's expectations could be achieved would be to use an institutional data repository for archive and publication of research data. Although a number of possible platforms for such a repository were available, all would have required some customisation both for research data and to support complete EPSRC compliance. As such, no institutional data repository had yet been selected for the University of Bath. Another facet of the Technical Data Coordinator's role was therefore to develop a specification for such a repository. It was essential that the specified repository would meet the needs of the EPSRC and other UK funding bodies, and also be usable by the university's researchers. In addition, it was intended that the specification would allow for future integration with existing research infrastructure, notably the open access publications repository, Opus,[16] and the Current Research Information System (CRIS).

To develop this specification, the Technical Data Coordinator data collated information from an institution-wide survey of researchers on data management issues, and conducted a series of one-to-one interviews with representatives from key stakeholder and advisory groups, including Computing Services, the Library, the Research Support Office and UKOLN. The Data Scientist was interviewed to represent the needs of the EPSRC and other external partners, such as publishers. The Technical Data Coordinator assembled a series of data repository user stories (Cope, 2013c) from which the specification was distilled and prioritised according to the

[15] More information about the pilot training workshop and the exercise testing data management planning tools is available via two Research360 blog posts: http://blogs.bath.ac.uk/research360/2012/02/rdm101-intro-definitions/ and http://blogs.bath.ac.uk/research360/2012/03/rdm101-data-management-planning/

[16] University of Bath Online Publications Store (Opus): http://opus.bath.ac.uk/

needs of the University and the Research360 project. This specification was used to commission a pilot data repository, developed by EPrints Services[17].

An example of how the specification met the requirements of the EPSRC's policy, researchers in the institution and their commercial collaborators was the requirement for an option to mandate input of the core DataCite[18] metadata fields. This would pave the way for minting of Digital Object Identifiers[19] (DOIs) for datasets in the future. This functionality would not only meet the EPSRC's fifth expectation[20] that digital data is issued with a 'robust digital object identifier' but it would also allow researchers to format citations for their data and include persistent identifiers in publications, thus promoting discovery, re-use and attribution of their data and increasing their research's impact. Similarly, the requirement that a basic metadata schema must include licensing and embargo periods would comply with the EPSRC's sixth expectation[21] relating to restricted access to commercially confidential data, and also allow adherence to collaboration agreements with commercial partners.

In addition to focusing on the technical aspects of data management, the Technical Data Coordinator's role was expanded to include provision of data management planning expertise. This involved development of a comprehensive suite of data management planning tools, including templates and guidance, with versions designed specifically for academic staff (Cope, 2013d; Cope, 2013e) and postgraduate research students (Cope, 2013a; Cope, 2013b).

It is anticipated that the Technical Data Coordinator's role will continue to develop once customisation of the institutional data repository commences post-project completion. While delivery of technical expertise, training and support will continue, the focus will increasingly be on technical development, with provision of data management planning tools becoming the responsibility of the Institutional Data Scientist.

Shared Responsibilities

Many of the activities of the Institutional Data Scientist and Technical Data Coordinator overlapped as they worked together to deliver training, draft technical reports, gather requirements and present at dissemination events. This close collaboration meant that the increasing workload of data management support could be shared. For example, both roles were able to provide support for individual

[17] EPrints Services: http://www.eprints.org/services/

[18] DataCite: http://www.datacite.org/

[19] Digital Object Identifiers: http://www.doi.org/

[20] EPSRC Expectation v: "Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the internet; in each case the metadata must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it. Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (For example, as available through the DataCite organisation - http://datacite.org)."

[21] EPSRC Expectation vi: "Where access to the data is restricted the published metadata should also give the reason and summarise the conditions which must be satisfied for access to be granted. For example 'commercially confidential' data, in which a business organisation has a legitimate interest, might be made available to others subject to a suitable legally enforceable non-disclosure agreement."

researchers, ensuring that requests for help submitted to research-data@bath.ac.uk were always responded to and resolved in a timely manner. It is important to note that half of the requests for support received by the project team originated outside the project's focal Faculties of Science and Engineering and, based on details provided by researchers, the majority of requests for help related to funders other than the EPSRC, predominantly the Economic and Social Research Council (ESRC), the Biotechnology and Biological Sciences Research Council (BBSRC), the Medical Research Council (MRC) and the National Health Service (NHS). This was because support requests were often the result of these funders requiring submission of a DMP as part of a grant application, something that is not currently required by the EPSRC. Although the Research360 project focused on the EPSRC's requirements, the number of requests for help with other funder policies raised the question of whether the project team should also support non-EPSRC-funded researchers.

Surveys of existing data management practice amongst University of Bath researchers (Jones, 2011; Pink, Cope and Jones, 2013) had highlighted considerable researcher uncertainty about data management and a lack of awareness of existing data infrastructure. Declining support to researchers who had requested help with other funders' policies, particularly when their other projects might be funded by the EPSRC, would have reinforced these misconceptions and the damaged reputation of data management would have spread rapidly amongst the University of Bath's close research community. It was therefore considered essential to support as many researchers as possible, regardless of their funding source, in order to enhance the status of data management and to initiate a general cultural change in how research data are managed. As a result, all researchers would be compliant with their funder's policies, including those whose research is supported by the EPSRC. This demonstrated an important lesson: that meeting the requirements of one particular funder cannot be achieved at the expense of another and that, perhaps surprisingly, these requirements can be met by instead focussing on the needs of the researchers, the institution and other external partners.

Use of the resources provided by both the Data Scientist and Technical Data Coordinator has, to date, been voluntary and dependent on researchers seeking out the assistance they require as and when a need arises. However, in order for the institution to achieve full compliance with the EPSRC's policy, it will be necessary to ensure that all researchers are aware not only of their responsibilities under the policy but also of the support the university offers to help them manage their data. One method of achieving this is advocacy, either directly or via word-of-mouth between researchers, giving rise to the cultural change previously described. However, this can take time and in order to meet the EPSRC's rapidly approaching 2015 deadline, the adoption of a more prescriptive method was required.

Policy Development

In order to ensure compliance with the EPSRC's third data expectation[22] relating to organisational policies and associated processes, it was necessary to develop a new, high-level policy for research data management. The development of this policy for the University of Bath typifies the challenge of reconciling conflicting internal and external drivers. Like many other UK universities, the University of Bath initially based its draft policy on the University of Edinburgh's influential Policy for Management of Research Data (2011). However, internal guidance on policy development quickly established that, since policies generally comprise requirements that are both measurable and enforceable, the policy ought not to consist of a purely aspirational set of principles. This change in style raised a number of questions: to whom and what would the policy apply, and how could compliance be achieved before a full data management infrastructure was in place?

When considering the scope of the policy, it was clear that at a very minimum it must cover all research funded by the EPSRC and, by extension, all research council-funded research. Immediately, the question of to whom the policy should apply became pertinent. University staff are contractually obliged to comply with relevant university policies. As many of the university's postgraduate research students are funded by the EPSRC and other research councils, it was essential that their research was also covered by the policy. As such, the policy team considered expanding the scope of the policy to include all research owned by the university. However, this caused difficulties in two areas. Firstly, it would have excluded dissertation projects undertaken by final-year undergraduate and taught postgraduate students. However, whilst the project team felt that experience of data management would be a valuable skill for graduates entering postgraduate research or employment, feedback from researchers was that mandating provision of data management training by policy could be considered excessive.

The second, more pressing problem was that the overlap between data ownership and data management was complicated by the collaborative nature of most research undertaken by the university. The nature of these research partnerships is defined by collaboration agreements, which tend to be complex legal documents between a number of different academic and commercial, national and international partners. Existing collaboration agreements generally do not explicitly reference research data as an output, let alone the long-term preservation and publication of such data. Further, in order to protect the commercial interests of industrial partners, it can sometimes be necessary for funding council policies to be flexibly interpreted, something that the funding councils tend to be amenable to in order to encourage these research partnerships. As such, developing a data management policy that simultaneously mandated compliance with funding council policies, whilst containing sufficient caveats to promote future collaborations with industry, proved extremely difficult. Feedback from researchers on early drafts of the policy suggested that

[22] EPSRC Expectation iii: "Each research organisation will have specific policies and associated processes to maintain effective internal awareness of their publicly-funded research data holdings and of requests by third parties to access such data; all of their researchers or research students funded by EPSRC will be required to comply with research organisation policies in this area or, in exceptional circumstances, to provide justification of why this is not possible."

inclusion of too many sub-clauses made the policy difficult to read, understand and subsequently comply with.

The policy development team therefore decided to separate data management from data ownership. An alternative solution was to apply the policy to all activities classed as 'research', as defined in the internationally accepted OECD Frascati Manual. Research was therefore defined as 'creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications' (OECD, 2002). The advantage of this definition was that it would include both research council-funded research and research funded by other funding bodies, such as charities and industry. However, it also included research that the university provides as a service for external researchers, such as the industrial consultancy provided by the Microscopy and Analysis Suite[23]. As the outputs of this research do not belong to the university, and the storage, retention and access to such data are generally defined by the contract, this type of research had to be explicitly excluded from the policy.

Another type of research that had to be excluded from the policy was that undertaken by postgraduate research students studying for an EngD or similar professional doctorate. These students, whilst still members of the university, conduct their research embedded entirely in external organisations and, as such, would use research infrastructure provided by their host organisation. The university would therefore be unable to mandate how research data created by these students were managed. Limitation of the policy's scope to research carried out at and for the university was therefore required.

The second aspect of policy development that proved difficult was allowing compliance before a full data management infrastructure was in place. In order to ensure that the university is fully compliant with the EPSRC's policy by 2015, it is necessary to mandate how data created by current projects are managed. For example, EPSRC-funded projects started since the publication of the nine expectations will have to ensure that research data 'be made as widely and freely available as possible' and that structured metadata be 'sufficient to allow others to understand…why, when and how it was generated.' For researchers publishing their work in 2015 it would be unfeasible, or in some cases impossible, for them to provide this information retrospectively about data created several years previously. The process of capturing descriptive metadata must therefore start as soon as possible in order for researchers to comply with the policy in the future. However, a research data survey carried out by the Research360 project team determined that some researchers (16.4% of 210 respondents) do not currently document their data sufficiently for others to understand them, and some perceived preparation of data for publication to be an additional burden on their time (Pink, Cope and Jones, 2013). Researchers are under a lot of competing pressures and without a policy mandating data documentation and publication, some researchers are unlikely to do this voluntarily.

This raises a further question about where researchers publish their data. Some disciplines, such as the biological sciences, structural chemistry and the social sciences, have a long history of open access data archiving, but for many researchers there are no national or internationally maintained data repositories where data can be

[23] University of Bath Microscopy and Analysis Suite: http://www.bath.ac.uk/facilities/mas/industry/

archived and published. This is particularly so within engineering and therefore of concern with regard to EPSRC-funded research. Use of an institutional data repository would enable researchers to fulfil many of the EPSRC's expectations, particularly relating to publication of structured metadata, use of access restrictions, secure preservation for ten years and use of persistent identifiers. However, by the end of the Research360 project this repository was still in the early stages of development and not anticipated to be fully customised and ready for use until after approval of the data policy.

As previously discussed, compliance with relevant university policies is mandatory for all staff. A policy that could not be complied with because the necessary infrastructure was not yet in place would immediately place all research staff in breach of that policy. To avoid this problem, other UK universities have used a number of approaches. For example, the University of Edinburgh included a statement which 'acknowledged that this is an aspirational policy, and that implementation will take some years.' Alternatively, the University of Bristol's draft policy consisted of a set of guiding principles[24] that sought to encourage researcher practice, rather than mandate activities. Due to the contractual nature of Bath polices, neither of these approaches was deemed suitable. Instead, it was agreed that the policy would be accompanied by an additional set of guidelines that would demonstrate to researchers how, in the interim period while the full data infrastructure is developed, minimal compliance could be achieved using resources already in place. For example, publication of the existence of data with details of how the data could be accessed, possibly via provision of a contact email address, could be achieved via the existing CRIS. An advantage of this solution was that separation of a high-level policy from more detailed guidance would allow the latter to be frequently updated as and when the data infrastructure is developed, without having to re-submit the policy itself for approval.

A final area of potential conflict between the various funders of academic research, as mentioned above, is the desire of publicly funded research councils to promote availability and access to research data, whereas many researchers and their commercial partners want to be able to restrict access to data. Such restriction allows them to maximise their return on the time and skill they have invested in creating the data by publishing a number of articles based on them, or sometimes by commercialising results. The EPSRC's policy, like that of many other funding bodies, recognises that there may be 'available exemptions which may be used, should the need arise, to justify the withholding of research data' and that these might include '...'commercially confidential' data, in which a business organisation has a legitimate interest.' Translating this into policy should, in theory, include a strong statement advocating publication of research data, whilst including caveats to allow for the withholding of data, where appropriate. In practice however, researchers need clear guidelines about what constitutes 'appropriate' or 'damage' to the research process due to inappropriate release of data. Does, for example, a competitor research group using published data for an identical follow-up study and then publishing the findings before the data's creators constitute damage to the research process? Until data publication is considered equal to article publication in terms of career development, many researchers might argue that it does. Correctly wording the policy to ensure that researchers do not all apply 'appropriate' caveats to publication of their data is likely

[24] University of Bristol Research Data Management Principles: http://data.bris.ac.uk/research/support/

to prove difficult and it may take time to determine how successfully the policy achieves the balance between funding council requirements and researcher interests.

Conclusion

The Research360 project concluded that for a research-intensive institution to achieve full compliance with a particular funder's policy, it can, perhaps counter-intuitively, be necessary to focus instead on fulfilling the needs of the institution, its external partners and researchers funded by other bodies. Development of the exemplar 'Roadmap for EPSRC' demonstrated the importance of aligning a new data management infrastructure with the existing strategic goals of the institution. To gain the support and resource required to implement this roadmap required exploration of how a range of stakeholders beyond the institution and focal funder would also benefit from investment in improved data management. Once high-level plans for infrastructure development are in place, it is the researchers themselves who must change how they manage their data to comply with their funders' policies. These researchers require support, not only in terms of technical infrastructure such as a repository for data archive and publication, but also in the form of guidance and individual assistance. The two data management roles described here have been pivotal in providing this support. Finally, that development of the institutional data policy extended beyond the extent of the Research360 project demonstrated how difficult it can be to reconcile the finer details of both funder and institutional needs, particularly before a supporting infrastructure has been fully implemented. Looking to the future, the continued provision of a research data service to support all researchers, regardless of their funding source, will continue the cultural change already started, meaning that best practice in data management will become 'business as normal' and all researchers will comply with their funders' policies, including the subset funded by the EPSRC.

Acknowledgements

The author wishes to thank: Jisc, for funding the Research360 project; members of the Research360 project team and the project's Steering Group; Charles Beagrie Ltd, who led development of the Business Case and supporting documents; many staff and researchers across the university, particularly students within the Doctoral Training Centre in Sustainable Chemical Technologies[25], who contributed time, expertise, support and guidance to the project. The author also thanks two reviewers for feedback on earlier drafts of this paper.

References

Beagrie, N., McKen, K. & Pink, C. (2012). *Research data management and REF2014.* Retrieved from http://opus.bath.ac.uk/35518/

Beagrie, N. & Pink, C. (2012). *Benefits from research data management in universities for industry and not-for-profit research partners.* Retrieved from http://opus.bath.ac.uk/32509/

[25] University of Bath Centre for Sustainable Chemical Technologies: http://www.bath.ac.uk/csct/

Beagrie, N., Chruszcz, J. & Lavoie, B. (2008). *Keeping research data safe: A cost model and guidance for UK universities*. Retrieved from http://sitecore.jisc.ac.uk/publications/reports/2008/keepingresearchdatasafe.aspx

Beitz, A., Dharmawardena, K. & Searle, S. (2012). *Monash University research data management strategy*. Report for the Research Data Management Subcommittee of the Monash Research Committee. Retrieved from http://monash.edu/library/researchdata/about/strategy/

Cope, J. (2013a). *Data management plan for postgraduate researchers.* Retrieved from http://opus.bath.ac.uk/30772/

Cope, J. (2013b). *Data management planning guidance for postgraduate researchers.* Retrieved from http://opus.bath.ac.uk/36009/

Cope, J. (2013c). *Institutional data repository user stories.* Retrieved from http://opus.bath.ac.uk/34082/

Cope, J. (2013d). *University of Bath data management plan template.* Retrieved from http://opus.bath.ac.uk/36360/

Cope, J. (2013e). *Writing your data management plan: University of Bath DMP template guidance.* Retrieved from http://opus.bath.ac.uk/36359/

Digital Curation Centre. (2012). *The essential data roadmap.* DCC Blog. Retrieved from http://www.dcc.ac.uk/news/essential-data-roadmap

Jones, K. (2011). *Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology at the University of Bath.* Retrieved from http://opus.bath.ac.uk/24960/

Jones, S. (2012). *Developing a roadmap for RDM.* DCC Blog. Retrieved from http://www.dcc.ac.uk/news/developing-roadmap-rdm

Lyon, L. & Pink, C. (2012). *University of Bath roadmap for EPSRC: Compliance with research data management expectations*. Retrieved from http://opus.bath.ac.uk/31279

Organisation for Economic Co-operation and Development. (2002). *Frascati manual: Proposed standard practice for surveys on research and experimental development*. Paris: Author.

Pink, C., Cope, J., Jordan, K. & Jones, K. (2013). *Research360: Faculty-Industry Data Requirements Report.* Retrieved from http://opus.bath.ac.uk/36361/

University of Edinburgh. (2011). Research data management policy. Retrieved from http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy