**University of Southampton**

**A Partnership Approach to Research Data Management**

**Mark L. Brown**

**Wendy White**

This is a preprint of a chapter accepted for publication by Facet Publishing. This extract has been taken from the author’s original manuscript and has not been edited. The definitive version of this piece may be found in Graham Pryor, Sarah Jones and Angus Whyte, *Delivering Research Data Management Services: Fundamentals of good practice*, 2013, Facet Publishing, which can be purchased from [www.facetpublishing.co.uk](http://www.facetpublishing.co.uk).  The author agrees not to update the preprint or replace it with the published version of the chapter. For terms of reuse see: <http://eprints.soton.ac.uk/id/eprint/356247>

By 2009 researchers at the University of Southampton had been working for some time with the issues around eScience and the challenges posed by integrating research data with publication.[[1]](#endnote-1) Individual researchers were already engaged in national collaborations, but it was the experience of collaboration around the UK Research Data Service (UKRDS) feasibility project in 2008-9 which acted as the catalyst for initiatives to support researchers across the institution in managing their research data.[[2]](#endnote-2) UKRDS was a sector wide initiative to investigate how the UK could respond to the increasing pressure on institutions to manage their researchers’ data. It initiated a corpus of work on the complexities of storage, retrieval, preservation and re-use, and if the proposal for a national framework did not go forward, the knowledge gained on the issues for the successful management of research data provided the background for the next phase of development.[[3]](#endnote-3) UKRDS was built on a partnership between the librarians, heads of computing services and leading researchers in the major research Universities who together with JISC and HEFCE set out a programme to respond to the needs of the research community. It was this collaborative approach which was taken forward in the next phase of development.

The University of Southampton is a major research-led University with a broad spread of disciplines and a recipient of a significant level of research income. In line with the major commitment to science and engineering, the University had since 2000 been investing considerable resource in high performance computing with the consequent increase in data output. In terms of managing research data, there were some disciplines for which the deposit of data for the purposes of archiving and sharing was well established, but for the majority of research areas there was no corresponding facility. UKRDS estimated that 21% of UK researchers used a national or international facility, and that there was an increasing level of data sharing. At Southampton this was reflected in the deposit of data in the national datacentres, such as those provided by NERC, ESRC, UKDA or Rutherford-Appleton,[[4]](#endnote-4) but there were important areas left without a model for data deposit and archiving and little provision to match the level of interdisciplinary and trans-institutional research being conducted. One of the aims of UKRDS had been to address the issue of archiving for those disciplines without a data centre, and there was disappointment that the proposal for a national approach did not go forward. The withdrawal of funding from the AHDS in 2008-9 also highlighted potential vulnerabilities for a University with a strong commitment to humanities.

Southampton has a long tradition of supporting open access for research outputs. The two core academic support services, the University Library and the University Computing Service (iSolutions), were partners in these developments, and it was a natural step to consider the role of data within the open access environment. The University had a strong track record of working on JISC funded projects and welcomed the positioning of Research Libraries UK, the Russell Group IT Directors Group and the JISC in pushing forward the agenda and providing opportunities to engage with initiatives at a national level. Within the University a community of shared interest evolved as a partnership between the two services and major research groups in Engineering, Archaeology, Computer Science and Chemistry.

***The Institutional Data Management Blueprint Project***

This community approach underpinned the partnership established through the Institutional Data Management Blueprint Project (IDMB), part-funded by JISC under the Managing Research Data (JISCMRD) programme, which ran from October 2009 to September 2011.[[5]](#endnote-5) The JISC programme dovetailed with the final report of UKRDS which appeared in May 2010. IDMB was seen by its supporters as a ‘great leap forward’ in terms of taking a researcher-led approach to translating research data management principles into effective practice deliverable at institutional level. To achieve this the project team, representing both academic and service champions,[[6]](#endnote-6) set out to engage the University institutionally to secure support for a ten year roadmap for data management, jointly ‘owned’ by the Research and Enterprise Advisory Group, responsible for research strategy, and the University Systems Board, responsible for systems strategy.[[7]](#endnote-7)

IDMB was intended to combine two approaches. A bottom-up approach based on researchers’ needs designed to broaden the adoption of good practice, and a top-down approach designed to set out the requirements for institutional policies and infrastructure. UKRDS had identified a great deal of work already underway in the UK, and the JISC programme was intended to extend these to foster awareness and promote good practice. UKRDS had also heightened awareness of the value of the existing data centres and given the wide discipline range within the University, IDMB was in a good position to benefit from the existing links with those provided by the Natural Environmental Research Council (NERC), the Economic and Social Science Research Council (ESRC), the UK Data Archive (UKDA) and the Archaeology Data Service (ADS). The project also sought to apply the tools being developed nationally by the Digital Curation Centre.[[8]](#endnote-8) This blend of national and local perspectives was important in the success of the project.

***A research-led approach***

From the beginning the team were determined to exploit existing good practice within the institution and to raise awareness of the implications of not managing research data. The serious fire which destroyed the laboratories of the Optoelectronics Research Centre in October 2005[[9]](#endnote-9) had already had an effect on the researcher community’s awareness of the vulnerability of the research record, and the debate surrounding around climate science data at the University of East Anglia sharpened perceptions of the role of Freedom of Information Requests in the debate on access to publicly funded research.[[10]](#endnote-10)

## To cement the importance of a research-led approach, the project team ran a ‘kick-off workshop’ in March 2010 which attracted around 40 attendees. The profile report identified what participants considered to be current issues of concern, long term aspirations and ‘quick wins’. This was followed up by a data management survey using a questionnaire and selected in depth interviews, and an audit using the DCC Assessing Institutional Digital Assets (AIDA) toolkit to benchmark current capability at department and institutional level. In exploring the toolkit the project team were referencing the wider work being funded by JISC in the Integrated Data Management Planning Toolkit & Support (IDMP) project intended to support the projects’ use of research data management planning tools.[[11]](#endnote-11)

## Taking a research-led and evidence based approach helped to formulate a set of preliminary conclusions as the basis of recommendations to the University. The key conclusions were:

* there was a need from researchers to share data, both locally and globally;
* data management was carried out on an ad-hoc basis in many cases;
* researchers’ demand for storage was significant, and outstripped supply;
* researchers resorted to their own best efforts in many cases to overcome lack of central support;
* backup practices were not consistent, with users seeking a higher level of support;
* researchers wanted to keep their data for a long time;
* data curation and preservation was poorly supported;
* Schools data management capabilities varied widely.

From these stemmed a number of institutional challenges. Although there was evidence of good practice, there was no coherent approach to data management across disciplines, and the current business model for curation and preservation was neither scalable nor sustainable in terms of future demands. The audit revealed that researchers were becoming more conscious of the requirements from funders and there was a need to consider issues around IPR, sharing data and the protection of the University’s digital assets. Elements of a service infrastructure were in place, but they lacked capacity and coherence; training and guidance was rudimentary.

***Institutional challenges***

Encouraging the University to adopt an institutional approach to developing policy and infrastructure faced potential barriers. The University has a very broad discipline spread and a strong culture of autonomy which did not fit easily with a centralised approach. On the other hand as an institution with a highly collaborative research culture many researchers were aware of debates over managing research data and Southampton’s long standing commitment to open access extended naturally into the realm of open data.[[12]](#endnote-12) Extensive work with research and learning repositories had orientated researchers to the principles and benefits of the central deposit of research outputs, but delivering an effective technical and service infrastructure was clearly beyond the capacity of individual academic units. The services therefore saw it as their role to explore the idea of a set of centrally delivered services which would be flexible and responsive to local needs. Here there were a number of assets within the pattern of existing practice. Although there was still a mixed economy in IT, there had already been a partial centralisation and the University had set up mechanisms for the central evaluation and procurement for IT investment. The Library service successfully runs as a centralised service, and had well embedded partnerships with academic groups across disciplines as well as a culture of collaboration with both iSolutions [[13]](#endnote-13) and Research and Innovation Services (RIS). The high profile of the UK research libraries in UKRDS and their position in many of the JISC funded projects emphasised the importance of the role of libraries in the data management landscape and the University Library emerged as a service lead.

From an institutional perspective funder policies were shifting perceptions of the significance of data management as a strategic issue. In the course of the project RCUK published guidance setting out seven core principles on data policy which highlighted the principle that publicly funded research data should be generally made as widely and freely as possible in a timely and responsible manner, though with appropriate safeguards for the inappropriate release of data.[[14]](#endnote-14) For the University as a significant holder of EPSRC funding the EPSRC framework on research data, which was under discussion during 2010 and which was adopted by the Council in March 2011, was particularly important.[[15]](#endnote-15) In contradistinction to practice by research councils such as NERC which provided a national data centre, EPSRC clearly placed responsibility for policy and compliance with research institutions.[[16]](#endnote-16) For the project team these developments posed the issue of achieving an appropriate balance between assuming voluntary adherence to good practice and an element of compliance based on institutional policy. Policy alone however would not be sufficient to create a change in the assumptions behind research practice in key areas. The audit provided evidence that researchers were open to new practice as long as it was researcher-led, integrated into research workflow, reflective of discipline distinctions and supported by advice and training. Clarity over policy and responsive service support were essential to gaining their commitment of incorporating good practice into workflow.

The roadmap was designed to set out a staged approach over a ten year period during which it was accepted that policy and technical opportunities would inevitably shift. The accompanying Blueprint was built around the concept of a multifunctional team which could bring together the knowledge and expertise of both professionals and researchers within a flexible technical and service framework. The message to the University was IDMB was not a set of solutions imposed on researchers, but a pragmatic and iterative process, flexible enough to meet the needs of a multi-disciplinary research institution. This implied an approach based on promoting cultural and political change.

Senior representatives of the University Executive Group (UEG)[[17]](#endnote-17) were involved with the project from the beginning as members of the project steering group. Reports to Senate were made through the Research and Enterprise Advisory Group (REAG) which included the Associate Deans (Research and Enterprise) and senior representatives from the research support services. This forum provided governance for the project and a forum for discussing future proposals. In terms of taking forward the principles in the Blueprint the project identified three priorities: the formulation and agreement of an institutional policy, an advocacy and training programme for the research community, and a strategy for the storage and security of data. Each of these issues posed significant issues.

***Shaping a Data Management Policy***

The team considered it important to link discussion of data management policy to the earlier debates on open access at Southampton. The work to build support for the institutional repository had helped shape researchers’ response to the central deposit of publications, and had created a momentum for making research outputs more visible, whilst recognising the concerns over the sharing of data. This experience of long-term engagement was seen as important in winning support for the data management policy, which needed to be seen as part of the wider policy framework for research and to appeal to researchers in terms of their needs. Data management, however, raises additional implications in terms of organisation, technology and resource. The policy had to set the context within which the institution could make decisions about investment and service support without introducing a set of requirements which would inevitably be perceived as an unjustifiable additional burden on research time.

In line with this thinking the team adopted a dual approach. At institutional level the policy sets out the University’s assumptions on roles and responsibilities, provides guidance on what is expected and sets the framework for decision-making and governance. It also provides a link to other institutional policies such as IPR Policy. For researchers it provides guidance on policy and governance and provides a framework for them to feel supported in responding to both internal priorities and external requirements. In terms of the researchers’ perspectives it sets a framework around the implications of changing funder mandates, the management of data workflow, the access, retrieval and security of data, and the facilitation of collaborative work with internal or external partners. By attaching ownership of the Blueprint jointly to the Research and Enterprise Advisory Group responsible for University research strategy and the University Systems Board responsible for overarching strategic decisions on University systems, it was intended that the policy would be embedded within University governance structures.[[18]](#endnote-18)

Having set out a draft policy it was important to put in place processes to support researchers in adhering to the base precepts. Researchers are working in a complex funder economy in which some research councils provide a national data centre and others assume institutional data management structures. In both cases, however, the issues of integrated workflow, data sharing, technical support and training for good practice would benefit from an institutional approach. The two priorities emerged as storage, security, curation and preservation of the data on the one hand, and advice and support to the researchers in managing their data on the other. Progress on the second element turned out to be easier than the first, not least because there was already a degree of good practice which could be used to extend support across disciplines.

***Integrated workflow: storage, security and archiving***

The Data Management Policy identifies the importance of the proper recording, maintenance, storage and security of research data, compliance with relevant regulations, including appropriate access and retrieval, and places the onus for achieving this on researchers. This reflected the consensus in the audit that individual researchers had to be responsible for management of their project data.

The choice made by researchers in their approach to storage was less consistent than might be assumed in the policies funders were requiring of institutions. In the audit researchers identified a wide array of data storage locations with 24% using their local computer, 34.9% using CD/DVD, USB flash drive or external hard disk, and only 24.3% using a file server either at the University or off-site.[[19]](#endnote-19) A significant number of respondents calculated that they held more than 100 GB of data, and 45.9% stated they kept their data forever. It was assumed from the return that a significant number of users managed this locally on a PC, CD/DVD, external hard drives or USB flash drives. More than 50% also indicated that they had experienced storage constraints, and overall the tracking of existing data was variable and often relied on paper logs.

When asked how the University could make data management and storage easier the main requirements cited were the need for more storage space, archiving, automated backup, security for sensitive data, a registry function, and integrated guidance and training. Some researchers in Electronics and Computer Science, Engineering Sciences and Archaeology also requested ‘EPrints for data’, reflecting the familiarity with the processes around the use of the research outputs repository. In terms of archiving only 10% deposited data with another external service, an average stretching between 28.6% in Psychology and Social Sciences and 9% in Geography.[[20]](#endnote-20)

This kaleidoscope of current practice revealed a potential mismatch between the way in which researchers were managing their data and the implications of the requirements from funders, most importantly for the University, the EPSRC.[[21]](#endnote-21) In the expectations set out by the EPSRC emphasis was given to the provision of ‘appropriately structured metadata describing the research data held which would be freely accessible on the internet, and 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’. These provisions made an assumption about institutional infrastructure and set a date, 1 April 2015, by which it was expected to be operable. One response to this was to put in place the policy; a second was to set in motion a debate on the deliverability of a suitable infrastructure.

In response to the debate on infrastructure the IDMB team produced an outline business model based on the assumption that the University would provide a secure and sustainable repository capable of hosting the University’s entire digital assets. Not surprisingly the University had no detailed knowledge of the quantity of research data held, nor of its likely growth over a specific timeframe, and the model had to be constructed on the basis of the indicative data from the audit and estimates from the current mid-scale research storage platform. The estimates indicated a current total of the order of 0.8 – 1.2 PB rising to between 11.2PB and 21.2 PB by 2016/17. The high level architectural design which was used to inform the cost model included three layers, an active storage layer, a metadata layer and an archive storage layer. The model also considered staffing and facilities support as well as operating costs, and made assumptions about the costs of storage over time. This allowed some scenarios to be constructed with options for the University to consider investment on a phased basis.[[22]](#endnote-22)

Although this work was speculative, it did provide a very useful baseline for assessing potential investment. This in turn had an effect on perceptions at senior University level of the pressures resulting from the combination of the growing size of institutional research data output and the requirements by funders for its management and accessibility. The knowledge and expertise being developed by the IDMB team was increasingly seen as an institutional asset. This was useful in the work being taken forward with designing a metadata layer for a registry function.

***Integrated workflow: metadata models***

In approaching this issue the IDMB team explored how a relatively straightforward metadata structure could be defined as a means of encouraging researchers to adopt external standards within an institution-wide registry. It was accepted that disciplines approached the value of metadata from different perspectives. Some used metadata just to identify and retrieve files whereas others, archaeology researchers for example, were accustomed to providing detailed metadata as part of their workflow and as preparation for deposit in the Archaeology Data Service.[[23]](#endnote-23) Some of the detailed analytical models at discipline level were seen to be too complex for researchers so the model was developed around a practical approach to encouraging researchers to submit baseline data.

The core metadata structure which evolved was based on Dublin Core, a standard already in use for data by the National Crystallography Centre at Southampton, [[24]](#endnote-24) and which built on the basic data deposit work undertaken in the DataShare project. [[25]](#endnote-25)The three level structure (Project, Discipline and Core) was devised to be as straightforward as possible and to be supported by usable tools for metadata assignment and import. Below is a figure showing the three level metadata structured applied by Archaeology, the pilot discipline, to a project exemplar:

 

The option for the inclusion of additional, more complex, metadata to be added as xml files at discipline specific level was also included in the specification, and work began on designing an ingest system. The initial scoping informed discussions over achieving an appropriate balance between a central, local and distributed technical infrastructure and was carried forward in the second JISC funded project, DataPool, which ran from October 2011 to March 2013.[[26]](#endnote-26)

***Integrated workflow: training and support***

In terms of training and support it was clear from the audit that researchers needed a wide ranging and flexible service model which would dovetail the separate services from the University Library, the Computing Service (iSolutions), Research and Innovation Services (RIS) and Legal Services. These services had a good track record of collaborative working and the experience of partnership in the IDMB project highlighted the value of closer integration. The team wanted to consider the need for support across the whole research cycle, encompassing the spectrum of research careers from the PhD student and the early career researchers to the mature research group engaged in largescale national and international collaborations.

As an initial pilot the IDMB team ran a training programme for archaeology PhD students. Training took the form of workshops where participants worked through a specific example looking at issues of storage and curation, and discussed potential solutions and the roles of different stakeholders in managing the data. Students were introduced to the three-layer metadata model and encouraged to think about how this would apply to their data collected in the course of fieldwork. Although this was a small pilot, it showed the value of integrating training into a broader programme of research skills training. The workshop approach was designed to test a template for implementation across other disciplines and from the feedback it was possible to determine priorities for the next phase of development which would incorporate deskside training and more tailored one-stop-shop guidance with further workshops.[[27]](#endnote-27) In the course of the pilot reference was also made to the other projects within the JISC programme to help reinforce the design.

***The impact of IDMB***

IDMB was intended to be the first phase of an iterative, dynamic model for supporting data management. The roadmap which emerged from the project identified three phases of development:

*Short-term (1-3 years)*

This phase was centred on the building of a core infrastructure, including an integrated approach to policy, technical infrastructure and support which could meet the demands of the growth in the level and complexity of research data, the requirements by funders and the need for the institution to manage its digital assets effectively.

The core components for this phase were:

* a robust institutional policy framework agreed and implemented by the institution;
* an agreed scalable and sustainable business model for storage based on the three components of active data, descriptive metadata and archive storage;
* a working institutional data repository which could satisfy researchers’ data management requirements for ingest, metadata creation and retrieval. It had to have sufficient capacity to attract users and offset the incentive to procure local solutions;
* a one-stop shop for data management advice and guidance to provide information on policy and legal issues to support the creation of data management plans, access to advice on technical capability, funder requirements and the benefits of managing data to exploit and share.

*Medium Term (3-6 years)*

During this phase it was assumed that the demands for the management of very large amounts of data of increased sophistication and complexity would increase, and that some disciplines would require potentially higher levels of data management input than can be managed within one institution. Although the cost of storage was likely to continue to decline, the management process itself would increase demand on staff skills. There would also be a higher profile for open and shared data and the value of pooling and sharing between institutions would be explored through specific exemplars.

The core components for this phase were:

* an extensible research information management framework to respond the variations in discipline needs;
* a comprehensive and affordable backup service for all based on the cost-benefits of backing up different classes of data;
* an effective data management repository model able to manage the potential full range of data deposit;
* building an infrastructure to respond to a commitment to open research data with a model for data publication;
* based on the cost-benefit analysis for backing up different classes of data, comprehensive solutions for managing research data across its whole lifecycle.;
* embedding data management training and support across the disciplines through partnership working between services and researchers;
* pilots with consortia to manage data collectively using standard infrastructure applications including cloud computing, and supported by shared staff knowledge and expertise.

***Longer term (6-10 years)***

Long-term aspirations would focus on providing significant benefits realisation across the whole University and a stable foundation for the future. The institution would have policies and infrastructure in place to make strategic judgements on how to manage its digital assets, and would have moved to a mixed-mode of data management within consortia or national frameworks. There would be a higher level of partnership between funders, organisations, local consortia and national facilities. Data management processes would be embedded throughout the research data lifecycle, and the infrastructure would fully support researchers with supply meeting demand via an easy-to-use data management service. This would significantly improve research productivity, allowing them to concentrate on their research, rather than worrying about data management logistics.

The core components for this phase were:

* coherent and flexible data management support across all disciplines across the whole data management lifecycle;
* agile business plans for continual improvement in response to changing requirements;
* commitment to innovation in open data publication and the infrastructure to support this across the institution;
* active participation in consortia and national framework agreements, contributing capacity and skills to building overall capability.

In promoting a debate over the role of policy and infrastructure in determining effective data management practice and institutional compliance with funder requirements, IDMB set the context for an institutional approach to research data management and defined the core elements in the framework for technological and organisational support. If compliance with funder requirements was a catalyst, it was not in itself a sufficient incentive for an institutional approach. The audit and the work undertaken with specific disciplines had engaged researchers in the issues underlying the management of their data; researchers had shown a very strong interest in adopting improved data management practice, and were open to working with the services to support their needs. It was clear however that any major investment would not be forthcoming without further evidence of value and impact.

***Initiating Phase 1 of the Blueprint: the DataPool project***

The JISC funded DataPool continuation project which ran from October 2011 to March 2013 provided the opportunity to progress and extend the first phase of that roadmap.[[28]](#endnote-28) Taking as a starting point the principles outlined in IDMB, the intention was to promote the framework across a full range of disciplines and to assess its adaptability to the more complex issues surrounding multi-disciplinary research.[[29]](#endnote-29) One of the challenges was to create sufficient impetus to engagement by both the researcher community and the University to take the issues beyond a project approach into a sustainable service infrastructure. IDMB had provided the framework and had established a network of senior managers, disciplinary leaders, faculty contacts and data producers; it was the role of DataPool to create the infrastructure needed to realise it. As with IDMB, DataPool was therefore as much about cultural as technical change

DataPool set out six key objectives designed to blend policy and infrastructure with local discipline perspectives. These were:

* implement the draft institutional research data management policy with an associated one-stop-shop of web guidance and data management planning advice;
* develop flexible support services and guidance for researchers extending across the research lifecycle;
* create and embed a range of training materials and workshops for postgraduates and early career researchers;
* enhance repository infrastructure to create comprehensive records of data outputs;
* scope options for storage and archiving including institutional structures, locally managed storage of small-scale outputs and a platform for sharing data;
* develop a suite of case studies to investigate multidisciplinary issues in depth including gathering granular evidence for cost analysis.

These strands were interdependent and were pursued in parallel to maximise the benefits of cross-fertilisation and to embed previous pilots into sustainable institutional services. The data management community was extended by embracing existing informal networks such as the multi-disciplinary University Strategic Research Groups (USRGs) and by engaging existing communication routes for research support between professional services and the academic community. The links with external sector providers were also extended through the involvement of senior academic co-investigators and the project Steering Group which had the benefit of advice from representatives from the University of Oxford, the British Oceanographic Data Centre, the DCC and the UK Data Archive.[[30]](#endnote-30)

***Sustaining the research-led approach***

Although compliance with funder policies had been a significant driver in engaging senior management, the team were conscious that researchers were most concerned with enhancing their research practice and heightening research impact. This posed issues for the team in terms of the variety of perspectives across the institution, and required close engagement with a wide range of researchers and nuanced disciplinary perspectives. This was particularly evident when drafts of the Research Data Management Policy were discussed. In the course of their liaison with colleagues, the nominated Faculty champions raised a variety of issues relating to the implementation of the policy across the data lifecycle, appraising data for long term retention, data security and sharing and the different roles and responsibilities in decision-making. Given the importance of open access for the University there were also discussions on the role of open data as a concept in data management practice.

In the light of these discussions the draft policy which had been drawn up under IDMB was modified and then presented to Senate in February 2012 with associated web guidance.[[31]](#endnote-31) It was emphasised in discussions at Senate that the policy was intended to be iterative and that the guidance would develop in response to feedback with relevant amendments being made on the basis of the experience. This was in part a response to the body of opinion which remained sceptical as to the financial implications and cost-benefit of implementing policy at an institutional level, and confirmed the team’s view that aside from funder requirements the emphasis should be on the embedding of good practice. The policy having been passed by Senate, the DataPool team were able to focus on developing the service infrastructure.[[32]](#endnote-32)

|  |
| --- |
| ***Integrated workflow: storage, security and archiving - Phase 2***  IDMB had highlighted the complexities facing researchers over storage and archiving. The DataPool project team recognised that without development in this area researcher engagement would be limited. The cost modelling undertaken as part of the IDMB set out options, but did not provide fully scoped business models which could convince the University to undertake largescale investment. Attention therefore focused on options for developing one of the infrastructure components, a registry function for the deposit and tracking of data. The archaeology pilot had confirmed that even where a discipline had access to a national repository for the archiving and sharing of data, there was a need for local infrastructure to deposit and describe data.  In approaching this issue the team understood that a ‘one size fits all’ approach was unlikely to be successful. From the audit it was clear that researchers had preferences about how they gathered and explored their data, and in some cases this required high levels of security which made them sceptical towards central, networked solutions. The point at which a researcher decides to deposit data in a registry would differ between disciplines; some researchers would welcome a workflow management system similar to that offered in a Virtual Research Environment (VRE) while others would only want to deposit final data at the end of the project. In terms of ingest the audit indicated that researchers in some disciplines had a preference for depositing data through the existing ePrints research repository, whereas in terms of longer term infrastructure developments ePrints might not be able to sustain storage across all disciplines. As the University was engaged in a pilot to evaluate SharePoint 2010, it was decided to explore both ePrints and Sharepoint as potential registry systems using the three layer metadata model. As part of this evaluation it was hoped that some data might be interfaced directly from other corporate systems thereby easing the burden on researchers in adding data.  The SharePoint application was ambitious. The first phase of a model was developed which provided researchers with a facility to deposit, manage and share data with colleagues during the life-cycle of a project, and potentially to export data to an external repository. With input from a group of PhD students and researchers in a range of disciplines an initial working demonstrator was designed, but further work was dependent on a University decision over the University SharePoint 2010 evaluation. Due to delays in this process further work on implementation has not proceeded as planned, and the business-case for the next phase is still being assessed. Sharepoint proved to be potentially flexible and to offer the prospect of importing relevant data from other corporate systems such as HR and Finance, but despite the success of the initial pilot, the level of knowledge and expertise needed to develop the software outstripped the resources of the project. The work with Sharepoint underlined the tension between embedding research data management requirements into a large-scale institutional strategic IT deployment and the flexibility to respond quickly to researcher feedback. It is clear from the JISC Programme as a whole that seamless technical workflow through the stages of the research lifecycle across all disciplines at scale is a sector-wide challenge.  Attention therefore turned to the option to extend the ePrints research repository to encompass the function of a registry for data linked primarily to research papers. Whereas Sharepoint was perceived as a possible ‘front door’ for both managing working data and final secure deposit, the ePrints model is being developed for deposit and access to research findings supporting published research. This approach meets the requirement of funders to identify how research findings can be accessed.[[33]](#endnote-33) ePrints has the advantage of being familiar to researchers across the institution, and has a community of common interest across the sector. Although ePrints is not able to handle the storage of ‘big data’, it can act as a registry and facilitate the deposit of data in significant discipline areas. In pursuing this option Southampton was also able to benefit from the work taken forward on metadata modelling by Essex in the ReCollect EPrints data app,[[34]](#endnote-34) and from the discussions between Essex, Southampton, Glasgow, Leeds and EPrints Services on standards and field mapping. There was agreement that local implementations should keep core metadata such as that required for DataCite and INSPIRE in common and to register a commitment to engage as an EPrints community on future developments in this area. The ReCollect app is now integrated into the live service at Southampton and we anticipate that use of this service will provide additional feedback to inform service enhancements.  Using ePrints to facilitate data management requirements showed the advantage of enhancing at relatively low cost an existing service with a broad user base. The “innovation to service” model was associated with a parallel development, the use of automated tools to support minting of DataCite DOIs which have been developed by Chemistry. Building on the existing Crystallography repository where there is already expertise in using DOIs to link to publications, this has created momentum to explore options for a multidisciplinary approach to the use of identifiers with DataCite. The work undertaken by Chemistry in relation to the LabTrove notebooks has also investigated potential granularity of DOI links. This includes modelling how a landing page might work within the dynamic notebook environment while at the same time providing a snapshot to support a publication.[[35]](#endnote-35)  As DataCite is specifically set up for data and has a growing researcher-led engagement and sector support through the British Library, there is a commitment to move to the DataCite Service. The disciplinary activity with Crystallography has evolved a generic app to automate DOI minting in EPrints. This is now being implemented in the central ePrints Soton repository and we are starting to explore policy implications. Whilst there is perceived institutional risk in assigning a DOI inappropriately, there is strong steer from senior researchers to take a pragmatic view to development, working with trusted frequent users and those with datasets underpinning publications  as early adopters, whilst less clear cut and more unusual cases are discussed.  ***Integrated workflow: multidisciplinary case studies***  If there were advantages in building on existing services and extending their application across disciplines, the DataPool case studies highlighted the need to think through the implications for specific areas where such a generic approach might not meet researcher needs.[[36]](#endnote-36) In the report on imaging, for example, it was confirmed that there existed extensive guidance on the effective management of raster and 3D data, but this was unevenly distributed across disciplines. It was also shown that there was insufficient entry level guidance and insufficient resources available to assist researchers in applying general principles to their own work.[[37]](#endnote-37) The Integrated Modelling of European Migration (IMEM)Database case study created a database and visualisation tool which explored the characteristics of probability distributions which could be applied across disciplines[[38]](#endnote-38) and the Tweeting study investigated various approaches to capturing and archiving tweets.[[39]](#endnote-39) These studies provided evidence of where service and institutional support could provide benefits to the researcher, and emphasised the potential complications facing researchers in providing evidence to funders of their impact and archiving strategies.  The case studies have revealed some relatively “quick wins” to promote change in practice for example the addition of raster and 3D equipment to the cross institutional EPSRC funded national equipment registry which is being set up to share resources across institutions. They have also provided a level of detail on storage requirements that can feed into business planning and create a narrative around investment and value. Also of value to business modelling was work on the potential of shared and third party services. Of particular significance was the development of an app linking EPrints to Arkivum’s A-Stor archiving service. This could be a component of a range of business models, including the potential to link with the DataCite implementation and the emerging service model for LabTrove electronic notebooks.[[40]](#endnote-40) The JISC programme has inspired much discussion of possible shared-services solutions and this is an area which we will be continuing to explore.  ***The Data Management Planning Service***  Evidence from the case studies was fed into a new Data Management Planning Service (DMP). Given the range and complexity of research proposals, providing a generic advice service for specific discipline requirements poses particular issues for the building of a trusted partnership. Our approach has been to provide web-based guidance to help interpret funders’ requirements, to offer deskside consultation and to refer specific issues to specialists in the discipline or data area. Raising awareness of the service across the University has been important as a means to bring together knowledge and expertise and provide the best advice. Key roles in this area are those of the faculty business relationship managers for the central IT services (iSolutions), academic liaison staff in the Library supporting specific disciplines, collaboration managers, bid managers and research support officers in Research and Innovation Services who due to their involvement with the bid process can contribute to specific areas such as intellectual property.  The DMP service is also proving important for engaging more experienced researchers who are time pressured and can be harder to reach. Awareness of policy requirements among this group can sometimes be limited and the need to complete a DMP is a spur to engagement. In the Faculties of Medicine, Health Sciences and Natural and Environmental Sciences this has led to requests for training for principal investigators on research data management plans, which has in turn provided examples for the training DataPool has been developing for PhD researchers. Researchers have sometimes started out looking for a template solution, but the feedback for the face-to-face support is a positive way of enhancing the programme and providing links across projects to such exemplars as business-cases for more specialist data management.  **Developing a training programme** |

IDMB had piloted an initial training template based on a workshop approach for archaeology research students. The DMP service has led to direct requests for specific DMP training for all staff from early career researchers to PIs.

The following diagram [[41]](#endnote-41) illustrates the approach taken to engage with the various groups involved with research data management from the new postgraduate researcher to the experienced principal investigator. It shows how events are channelled through existing structures reflecting lifecycle, format and audience with co-delivery in each case between professional services and researchers.

In developing the training programme reference is being made to the VITAE Researcher Development Framework and parallel work in other JISC programme projects and the DCC which is broadening understanding of the way in which training can play an important role in supporting researchers. Co-delivery of training has been very effective particularly in sessions aimed at postgraduate or early career researchers, and PhD students have been involved in developing the supporting guidance, disciplinary case studies as well as co-delivering training.

Implementing the programme has revealed a number of important issues in achieving a balance between generic and specialised content, the position of the audience in terms of their knowledge and experience, and the appropriate length and format of the workshops. For generic sessions, for example, feedback indicates that it would be productive to provide more time for discussion on different aspects of data management to enhance thinking of how practices might be changed. It would also be useful to signpost level and content more clearly so that researchers gain the most from their attendance. Extending the range of training in terms of the research lifecycle, there is now a pilot to offer training at masters’ level; from October 2013 a new MSc Instrumental Analytical Chemistry which has been set up to embed data management as one of the modules.

An evolving programme of training for professional services staff has been designed which will be further developed by incorporating the results of the survey exploring levels of staff confidence in various research data management areas and knowledge of referral options. The survey has also been completed by staff in Oxford which provides useful comparative data and opportunities for joint/ regional training.[[42]](#endnote-42) This partnership between services and academic groups who are often focused on quite specific issues for their research is particularly valuable for engaging with the issues of multidisciplinary research, but it also raises issues about the breadth of knowledge and expertise which can be accrued by those offering training within the professional support services. The current survey to assess the needs of professional service staff in the Library and Computing and how these can be supported will facilitate further reflection and iteration. As the level of engagement by researchers and their expectations rise, there can be significant issues for the services in being able to respond effectively at the different levels required. The investigation will therefore be reviewing both the capacity and the scope for using a wider variety of approaches.

The DataPool project came to an end in March 2013. In terms of building institutional capacity within the framework laid out in the first phase of the IDMB roadmap, it had signalled some of the major challenges and showed how some solutions might be put in place. It had taken these forward to the next stage of development and significantly extended institutional and researcher engagement. Of the four key components outlined in the roadmap, three had been delivered to a point where they were poised to be recognised in policy and practice. An institutional policy framework had been agreed and implemented, the basis for working institutional data registry had been scoped within ePrints, and a one-stop shop for data management advice and guidance to provide information on policy, legal issues and guidance, had been put in place. The gap remains the development of an agreed scalable and sustainable business model for storage based on the three components of active data, descriptive metadata and archive storage. Scoping work had taken place, but as this developed it was clear that the financial and systems investment would be considerable. The University recognises the issues involved, but has to be convinced that a single, standalone institutional solution can be realistically funded and implemented. At the point of writing thinking is now pointing towards collaborative shared services.

**Reflections on current progress**

The IDMB and DataPool projects provided a focus for the University to address the question of how Southampton should design and deliver its research data management strategy. If some of what was presented in the IDMB roadmap was too far in the future to be embraced with certainty, the concept of a gradual and iterative process embodying good practice was established in institutional thinking and some key milestones defined. The progress made under IDMB set the scene for the programme of work in DataPool which began the process of embedding policy into services which could support researchers. The continuity between the two phases was reinforced by the existence of a common core project team who acted as change agents. Both projects were essentially about cultural change which was to be nurtured through a combination of a top down and a bottom up approach.

In terms of the top down element there is no doubt that a formal University policy on research data management was very important in achieving visibility for a collective approach across the University. The team was careful to work through a process of consultation before the policy was formally put to Senate, and that process made the policy more acceptable in the eyes of the research community. The team was also aware that without a responsive approach to some of the key concerns of researchers it would not be an effective mechanism to take forward the agenda. Through the common governance structure for both IDMB and DataPool, the DVC Research and Enterprise, the Provost and the Associate Deans (Research) were all associated with the investigations and outcomes. Senior University Management was therefore well informed about the issues, and the same governance model is being taken forward in the post-project phase, emphasising the long-term approach set out under IDMB. The team hopes that this will act as the political mainspring for future infrastructure investment; up until now convincing evidence for a return on investment is unclear. The next phase is to work alongside specific externally funded projects with each Faculty to strengthen disciplinary narratives and provide more granular evidence.

Although top down policy and support was important, the development of low cost, low overhead solutions to some of the challenges facing researchers as a result of institutional or funder requirements was a major incentive for cultural change. The researcher-led approach was taken forward in a number of ways. The project group itself was forged through a collaboration between academics across a variety of disciplines and the leading academic support services. PhD students and research fellows have been significant contributors, offering knowledge of specialist data needs and contributing to training. This peer engagement has meant that they have also acted as key change agents, often bridging gaps between research groups across disciplines. They have led project activity through the case studies, contributed to testing the SharePoint and ePrints developments, designed training material and led workshops often in conjunction with colleagues from the services. They have also contributed specific technical expertise as part of service development. For PhD and research fellows the informal networks have played a key role in developing and engaging with communities of practice.

PhD researchers, research fellows and technical experts have led strands of the project that aim to promote an ‘innovation to service’ approach that ties in with the mid-phase of the institutional roadmap. The multi-disciplinary case studies on imaging requirements and data visualisation for impact both provided evidence of innovation that can be embedded in services. The technical work which has provided EPrints bazaar apps to automate DataCite DOI minting and link EPrints to Arkivum storage gives us an opportunity for community implementation and service refinement. Further proof of concept work with DOI minting for Labtrove electronic notebooks and data transfer to Arkivum pave the way for possible shared services approaches with roles for third party providers.

Reflecting on the progress since 2009 it is clear that the issues surrounding research data management are becoming more complex rather than less. We now understand much more about the range of data to be managed, its size and sophistication and the expectations of researchers to manage workflows and share data. We also know that at institutional level the requirements of government and funders are placing potentially significant financial costs on institutions which they are finding challenging to discharge in the present financial climate. Our approach has been to build a partnership around discipline needs, researcher workflow, low cost technological applications and training support, and this will be integral to the way in which we continue with the implementation of the IDMB roadmap.

MLB.mlb.draft7.9June2013

1. Rachel Heery, Monica Duke, Michael Day, Liz Lyon), Michael B. Hursthouse, Jeremy G. Frey, Simon J. Coles, Christopher Gutteridge, Leslie A. Carr (2004), *Integrating Research Data into the Publication Workflow: the eBank UK experience,* University of Southampton, <http://eprints.soton.ac.uk/9705/1/PV2004-heery.pdf>. [↑](#endnote-ref-1)
2. UKRDS was in part a response to the 2007 report from the Office of Science and Innovation (OSI),

   Developing the UK’s e-infrastructure for science and innovation. (<http://www.nesc.ac.uk/documents/OSI/report.pdf>). UKRDS was funded by HEFCE under the shared services programme [↑](#endnote-ref-2)
3. UKRDS was perceived as a centre of excellence, as a standards-guiding body and as a source of expertise and information about data management and repositories, commissioning additional

   capacity and building on current best practice and facilities. <http://www.ukrds.ac.uk/>. [↑](#endnote-ref-3)
4. Only 12% of researchers surveyed did not make their data available in any way. UKRDS Final Proposal, p.13. <http://www.ukrds.ac.uk/resources/>. [↑](#endnote-ref-4)
5. <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>. [↑](#endnote-ref-5)
6. **Kenji Takeda (Engineering Sciences), Mark Brown (University Librarian), Wendy White, Head of Scholarly Communication, Simon Coles (Chemistry), Les Carr (Electronics and Computer Science, Eprints), Jeremy Frey (Chemistry), Peter Hancock (iSolutions), Graeme Earl (Archaeology).** [↑](#endnote-ref-6)
7. <http://www.southamptondata.org/>. [↑](#endnote-ref-7)
8. <http://www.dcc.ac.uk/>. The DCC has a a focus on building capacity, capability and skills for research data management across the UK's higher education research community, and was a key partner in the JISC funded projects. [↑](#endnote-ref-8)
9. <http://news.bbc.co.uk/1/hi/england/hampshire/4390048.stm>. [↑](#endnote-ref-9)
10. <http://www.timeshighereducation.co.uk/414350.article>. [↑](#endnote-ref-10)
11. This project ran from January 2010 to March 2011, <http://www.jisc.ac.uk/whatwedo/programmes/mrd/supportprojects/idmpsupport.aspx>. [↑](#endnote-ref-11)
12. See for example the early work of Jeremy Frey in Chemistry at Southampton, <http://eprints.soton.ac.uk/38009/2/www2006_panel.pdf>. [↑](#endnote-ref-12)
13. iSolutions is the name given to the University of Southampton IT Service. [↑](#endnote-ref-13)
14. <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>. [↑](#endnote-ref-14)
15. <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>. The DCC provides a directory of funder policies, <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>.

    [↑](#endnote-ref-15)
16. Section iii states ‘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.’ [↑](#endnote-ref-16)
17. Specifically the Deputy Vice Chancellor for Research and Enterprise, the Provost and the Registrar and Chief Operating Officer. [↑](#endnote-ref-17)
18. The Research Data Management Policy was agreed by Senate in February 2012 and published in the University Calendar <http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html>. [↑](#endnote-ref-18)
19. Data management questionnaire results: IDMB Project. The disciplines represented in the audit were Electronics and Computer Science, Engineering Sciences, Geography, Humanities, Maths, Psychology and Social Sciences. <http://eprints.soton.ac.uk/196243/1/IDMB_Survey_Report.pdf>. [↑](#endnote-ref-19)
20. There were a number of significant deposits outside the framework of the Research Councils data centres including the Africa Centre for Health and Population Studies, the Max Planck Institute for Demographic Research and TPTP (Thousands of Problems for Theorem Provers). [↑](#endnote-ref-20)
21. In the EPSRC policy there were an expectation that there would be ‘appropriately structured metadata describing the research data held which would be freely accessible on the internet, and 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.’ <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>. [↑](#endnote-ref-21)
22. The Business Model forms Appendix B of the IDBM itself <http://eprints.soton.ac.uk/196241/>. [↑](#endnote-ref-22)
23. <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors>. [↑](#endnote-ref-23)
24. <http://ecrystals.chem.soton.ac.uk/>. [↑](#endnote-ref-24)
25. <http://www.jisc.ac.uk/whatwedo/programmes/reppres/sue/datashare.aspx>. [↑](#endnote-ref-25)
26. <http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure/datapool.aspx>. [↑](#endnote-ref-26)
27. Earl, Graeme, White, Wendy and Wake, Pam (2011) IDMB Archaeology Case Study: Summary., University of Southampton, <http://eprints.soton.ac.uk/196237/>. [↑](#endnote-ref-27)
28. The DataPool website contains links to the reports and the blog maintained by Steve Hitchcock, joint project manager, throughout the project, <http://datapool.soton.ac.uk/>. [↑](#endnote-ref-28)
29. The DataPool team was the same as that for IDMB except that with the departure of Kenji Takeda the role of Principal Investigator was placed with the Library. [↑](#endnote-ref-29)
30. Steering Group: Louise Corti (Associate Director, UK Data Archive), Graham Pryor (Associate Director, Digital Curation Centre), Sally Rumsey (Digital Research Manager at The Bodleian Libraries, University of Oxford), Helen Snaith (National Oceanography Centre, Southampton), Philip Nelson (Pro-Vice Chancellor Research), Adam Wheeler (Provost and Deputy Vice Chancellor),  Mylene Ployart (Associate Director, Research and Innovation Services) and the project investigators. [↑](#endnote-ref-30)
31. The policy was published in the University Calendar with links to four guidance sheets on topics which the research community had prioritised in the discussions on the policy: Storage Options, Guidance on Retention Periods, Recommended Practices for Destruction of Data and Restricting Access to Research Data. http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html. [↑](#endnote-ref-31)
32. White, Wendy and Brown, Mark (2013) DataPool: Engaging with our Research Data Management Policy, University of Southampton, <http://eprints.soton.ac.uk/id/eprint/351945>. [↑](#endnote-ref-32)
33. This requirement is contained within the RCUK open access policy released in 2012e. The current guidance confirms that for ach publication must include a statement on access to the underlying research materials, <http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicy.pdf>.

    [↑](#endnote-ref-33)
34. See <http://www.data-archive.ac.uk/media/395364/rde_march2013_repositoryoutputs.pdf>. [↑](#endnote-ref-34)
35. <http://datapool.soton.ac.uk/2012/12/18/trialling-datacite-for-chemistry-lab-notebooks-and-repository-data-services/>. [↑](#endnote-ref-35)
36. Datapool carried out three case studies based around multidisciplinary imaging, data visualisation and impact, and tweeting and produced a cost-benefit analysis built around the management of significant sized data and the value of an institutional approach. This report included the Open Data Service at Southampton.

    Hitchcock, Steve (2013) *Cost benefit analysis experience of Southampton research data producers*<http://datapool.soton.ac.uk/2013/03/21/cost-benefit-analysis-experience-of-southampton-research-data-producers/>. [↑](#endnote-ref-36)
37. Beale, Gareth and Pagi, Hembo (2013) DataPool Imaging Case Study: Final Report. University of Southampton, <http://eprints.soton.ac.uk/id/eprint/350738>. [↑](#endnote-ref-37)
38. Wisniowski, Arkadiusz, Chivers, Martin and Whitton, Michael (2013) Integrated Modelling of European Migration Database Case Study. University of Southampton, <http://eprints.soton.ac.uk/id/eprint/350672>. [↑](#endnote-ref-38)
39. Hitchcock, Steve (2013) Collecting and archiving tweets: a DataPool case study, University of Southampton, <http://eprints.soton.ac.uk/id/eprint/350646>. [↑](#endnote-ref-39)
40. Arkivum A-Stor Storage Backend plugin,  <http://bazaar.eprints.org/285/> . [↑](#endnote-ref-40)
41. Byatt, Dorothy, Scott, Mark, Beale, Gareth, Cox, Simon J. and White, Wendy (2013) Developing Researcher Skills in Research Data Management: Training for the Future – a DataPool project report, University of Southampton, <http://eprints.soton.ac.uk/id/eprint/351026> [↑](#endnote-ref-41)
42. Byatt, Dorothy, De Luca, Federico, Gibbs, Harry, Patrick, Meriel, Rumsey, Sally and White, Wendy (2013), *Supporting Researchers with their Research Data Management: Professional Service Training Requirements – a DataPool project report*, University of Southampton, <http://eprints.soton.ac.uk/352107/>. [↑](#endnote-ref-42)