



Project StORe

WORK PACKAGE 5: EVALUATION

A final report to the Project Board on the process and outcomes of the project, including recommendations for further development

Graham Pryor
Project Manager, Source to Output Repositories

August 2007

JISC

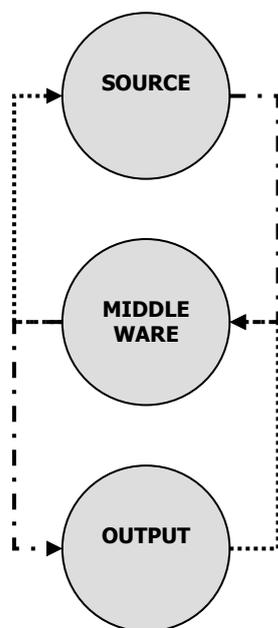


Contents

	Page
1 Introduction, scope and synthesis	5
2 Context and key outcomes from the survey of researchers	9
3 The pilot demonstrator – an assessment	13
4 Options for development and deployment	15
5 Afterword: the project management experience	17
6 References	21

1. Introduction, scope and synthesis

Project StORe was conceived by members of the UK's research library community as an initiative that would apply digital library technologies to the creation of new value for published research. Ostensibly a technical project within the JISC Digital Repositories Programme, 2005 - 2007¹, its primary objective has been the design of middleware to provide bi-directional links between *source*² repositories containing research data and *output*³ repositories containing research publications, with the aim of enabling direct electronic navigation between research papers and their source data. Using the StORe middleware, it was envisaged that researchers would be able



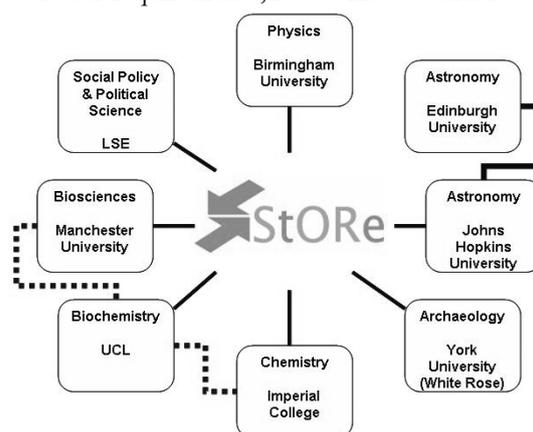
to relocate directly from within an electronic article to the source or synthesised data from which that article was derived and, reciprocally, direct access would be provided from source data to the publications associated with those data. A principal output from the project was to be an independently evaluated pilot demonstrator of the middleware.

In the earlier stages of the project it became apparent that the anticipated technical challenge of developing this new suite of functional middleware would be compounded by the diversity of the research data environment. This diversity was evident in the expression of cultural preferences and research traditions as well as from more practical aspects, such as the variety of research methodologies used. It was represented as much by tangible elements - for example, the range of scientific data outputs being deposited in *source* repositories - as by the broad spectrum of attitudes held within and between individual scientific domains towards data sharing and dissemination.

The relevance of traditional library skills and experience to the design of middleware for bridging across this overtly heterogeneous research environment had been taken as read, since librarians and information scientists had already been successful in the articulation of generic tools for use in other equally diverse contexts (e.g. metasearch interfaces to publisher and local databases, metadata harvesters and link resolvers, all based upon recent digital library protocols and standards such as OAI-PMH⁴ and qualified Dublin Core⁵); but in order to incorporate high quality contextual information in the middleware that would reflect actual user requirements, it was first essential to survey researcher behaviours and the processes employed in the generation, organisation and sharing of data. This survey of seven scientific domains, employing an online questionnaire and one-to-one interviews, was conducted in the Spring of 2006 by a team based at institutions in the UK and USA.

During the survey, researchers had reacted favourably to the opportunities anticipated from the projected StORe middleware, with 85% of those who responded predicting that the bi-directional function would prove advantageous. At the same time, whilst the initial plan to enhance the functionality of repositories had been laid to improve opportunities for information discovery and data curation, it became quickly evident that by promoting access to research data, the project was also challenging the familiar and traditional concept of academic publishing. This proved to be a source of disquiet amongst some of the researchers surveyed, most particularly on the subject of peer review.

Making public (or *publishing*) data that has not been through a rigorous and formal process of peer review would not be without risk, even with the application of such measures as embargoes



to govern the timely release of data, but it was encouraging to find amongst a periodically sceptical scientific community some affirmation of our view that the actual process of making data public can contribute to the validation of the data; even serving to protect the integrity of a scholarly article based upon it. The premise here is that where its underlying data is visible and when members of the research community can access and ‘peer review’ it unconstrained, the quality of the arguments made within any published paper should as a consequence be of the highest quality. Nonetheless, the StORe survey found that the cultural changes suggested by this recent option to ‘publish’ research data may still be difficult to achieve on account of a) a serious necessity to improve upon basic data management practices and b) the need to resolve compelling issues pertaining to data ownership.

Despite the barriers to cultural change encountered by the survey team, other influences contemporaneous to the project were found to be increasing the pressure on researchers to make their data available. In the UK, this pressure was evident from the research councils’ increasing concern that data generated from publicly funded research should be managed in ways that better reflect their value. Mid-way through Project StORe, in 2007, the BBSRC and MRC⁶ issued new data policies, both based on principles adopted from the OECD report on *Promoting Access to Public Research Data for Scientific, Economic and Social Development*⁷, which recognise that publicly funded research data are a public good, produced in the public interest, and should be openly available to the maximum extent possible. Recognising issues of ownership and intellectual property rights, both Councils would allow researchers a limited period of exclusive use, but they require that new research data must be properly curated throughout the information lifecycle and, when released, should include high quality contextual information, or metadata.

Albeit a policy framework has begun to emerge, the overall tenor of responses to the StORe survey suggest that its impact will be slow to resonate. As recently as June 2007 Liz Lyon’s consultancy report *Dealing with Data*⁸ concluded that when considering “the human aspects of data curation activities, many researchers appear to be unaware of the range of issues associated with data management best practice and there is a growing requirement for co-ordinated advocacy, training and skills programmes to equip the research community with the appropriate competencies”. Recognising that institutional repositories tend to be managed by ‘generalists’ located in libraries and information services, whereas subject repositories are normally staffed by domain experts with well-developed data handling skills, Dr Lyon recommends closer partnerships between institutional and subject repository staff as a prerequisite to building effective engagement between the broader repository movement and the research community.

A focus on the need for such engagement had informed the rationale for the StORe middleware since project inception. The generic specification developed by staff at the UK Data Archive was based not only on the set of repository enhancements identified from the survey of researchers, but also reflected key aspects of the cultural environment that governed approaches to networking and data sharing, privacy and protectionism, and self-reliance. Consequently, the model for bi-directional linkage emerged as a Web 2.0 design, similar to existing FOAF (Friend Of A Friend) services such as Flickr or MySpace and representing a familiar digital context for many researchers. This design allowed them to remain in control of their research output, leaving them to determine which of their data items should be selected for deposit and publication, whether they should be designated public or private items and who should be permitted to join their data collaborations as colleagues (or ‘friends’).

Project StORe has taken a particularly practical approach to developing data/publication links, based on the establishment of a simple but unique and persistent identifier for each item in a collection, and made at the point of ingest to a source repository. Other projects within the JISC digital repositories programme have pursued alternative solutions, some concentrating on standards for data citation, others focusing on the data lifecycle requirements of a single domain.

The independent evaluation of the StORe pilot demonstrator (Phase 1) by the National Centre for eSocial Science found that designing the system around a platform of research collaborations proved particularly appealing to users who tested the system. This result augurs well for development of the middleware as a workspace environment, an even more significant outcome

than the provision of a tool to enable traffic between data and publication repositories. With that endorsement, it is anticipated that the enhancements planned for 2007 - 2008 (described on page 15) will confirm the StORe middleware as a desirable component within a larger suite of data management solutions.

The following sections of this report each commence with reference to corresponding documents from the project's individual work packages. Their content has not been reproduced here but is freely available from the University of Edinburgh Research Archive, at www.era.lib.ed.ac.uk/, where all Project StORe documentation has been deposited and identified as the *StORe project* collection.

2. Context and key outcomes from the survey of researchers

<u>Work package reference</u>	<u>Documents produced</u>
2: Survey of Researcher Use of Repositories	Survey Report Part 1: cross-discipline report, August 2006 Survey Report Part 2: discipline reports - The Source to Output Repositories Project: Archaeology , July 2006 - Project StORe: Astronomy Report, August 2006 - Project StORe (Source-to-Output Repositories) Report for Biosciences , September 2006 - Project StORe (Source-to-Output Repositories) Report for Biosciences, Errata , October 2006 - Source to Output Repositories Project: Chemistry , August 2006 - Project StORe: Physics Report, August 2006 - The Source to Output Repository Project: Social Science , October 2006

The Project StORe survey was conducted between February and July 2006, using an online questionnaire followed by a series of structured interviews. More than 3,700 researchers were invited to complete the questionnaire, producing a response rate of 377 (>10%). The schedule of interviews, which were undertaken face-to-face, by telephone and in one instance as a group discussion, involved a minimum of 10 researchers in each of the seven disciplines surveyed. The cohort of researchers selected for interview was not limited to respondents to the questionnaire.

It was planned that subsequent to the survey a business analysis would determine from the questionnaire and interview data which critical functions were to be met by the StORe middleware. When structuring the survey, the requirements of the technical landscape had been uppermost, with particular interest being paid to the nature of workflows, any norms for repository use, the common attributes of data and research methods across disciplines, the functional enhancements to repositories that were perceived to be most desirable, and the characteristics of problems reported with existing repositories. However, the survey opened up a far broader territory than had originally been envisaged, exposing a range of cultural and behavioural features in the scientific research community that impacted radically on the project's technical rationale. Several aspects of this insight were subsequently corroborated by other studies, notably the Lyon report referred to above⁹, as well by the Research Information Network's UK survey of 400 researchers and 50 librarians¹⁰, thereby increasing the confidence of StORe project staff, who had endeavoured to design the middleware around the research culture's more insistent facets.

The documentary outputs from the survey comprised individual reports from six of the seven disciplines surveyed, plus a cross-discipline report. No report from biochemistry was delivered, although sufficient analytical data was provided from the biochemistry survey to afford its inclusion in the business analysis. The cross-discipline report contains

- a description of the purpose, context and method of the survey
- a summary of the behavioural evidence gathered
- a commentary on responses made to the questionnaire
- an introduction to the interviews
- structured responses to the questionnaire
- free text responses to the questionnaire
- identification of the participating institutions and disciplines surveyed
- a list of the source and output repositories nominated for comment and evaluation during the survey.

Key messages from the survey

As is emphasised in the cross-discipline report, the diversity of responses to the survey dictates that the testimony summarised here can only represent themes tending towards consensus.

Research culture

- The traditional image of self-sufficiency and independence continues to be a significant feature of researchers/research groups
- Networking is conducted principally on an informal and personal basis, with printed or electronic mail accounting for more than 50% of data sharing
- A disinclination to seek or utilise support services in the management and use of research outputs is pervasive, despite the expertise of information professionals being recognised in specialist areas such as the design and maintenance of metadata
- Many researchers describe a genuine dependency toward the ownership and safekeeping of their research output, one that produces a demand for caution with respect to its release, although this varies between disciplines (astronomy being the most relaxed in this respect)

Repositories – awareness and use

- A limited awareness and understanding of repositories was demonstrated in each of the seven disciplines, despite the existence of well established source and output repositories
- The most considerable evidence of sophistication was provided by astronomy, the biosciences and physics, and in archaeology there is a high awareness of the value of the discipline's principal source repository (*Archaeology Data Service*)
- Proof of provenance, ease of deposit and use, and comprehensiveness of content are regarded as key to the acceptance of repositories as credible resources
- Institutional repositories are not favoured as a means of storage or dissemination for research, neither are they a first resort when seeking resources
- Globally provided data resources, organised according to discipline, are preferred
- In the case of published articles, the use of electronic journals remains the most valued option for dissemination and retrieval
- A majority of research postgraduates are not receiving structured exposure to or training in the use of repositories, nor in good data management practice

Data management, access and sharing

- The principle of open access to publicly funded research outputs is universally supported, although this position can oscillate according to a researcher's role either as producer or consumer
- Commonly, inadequate data management techniques are applied to the protection and curation of research outputs
- Whilst the importance of good metadata is recognised by most researchers, the use of standard schemes is limited
- There is serious concern that the provision of easy access to source data would introduce a damaging risk of predation
- The value of raw data to researchers outwith the originating team/programme is to be questioned and the criticality of attaching interpretative metadata emphasised
- A simple *Google* type approach to searching is often preferred over methods offering greater specificity, particularly where Web based research replaces browsing in a library

Conclusion

One could easily be tempted by several of these statements to reach a negative view of the scientific community's position with respect to repositories, even towards data management in general. Such a view would be erroneous. The balance of responses to the StORe survey demonstrated a high level of familiarity with the principles of best practice for managing data and, when invited, a wealth of informed suggestions were given for improving the functionality of available and projected repository services, which supported a more positive evaluation. Upon reflection, the messages that at first had seemed simply to convey a negative outlook on the part of respondents appeared instead to be evidence of the failure of advocacy campaigns to promote an enduring awareness of repositories, the impotence of strategies to encourage their use, and the absence of a partnership between the practitioners of eScience and those providing the infrastructure and standards that support it. Explaining why these conditions persist was beyond the remit of the project, but it was clear that for a large proportion of the population surveyed, repositories had not by their presence alone persuaded researchers to recognise their value.

In a their second author study on open access¹¹, Alma Swan and Sheridan Brown found that over 70% of researchers use *Google* or *Google Scholar* as their first-choice tool for finding information on a topic, and the StORe survey came upon repeated references to the need for repositories to employ a *Google* type search engine. One might then claim that *Google* has defined the information landscape for science researchers requiring quick and easy retrieval of information. Furthermore, when seeking information from scholarly articles, publisher sites occupy the centre ground as the natural source. Repositories cannot be developed in isolation from this *status quo*, and their relevance to popular practice will need to be demonstrated through meeting or orientation around the broader demands for improved functionality, as was intimated when researchers were invited to suggest how repositories might be improved¹²:

- “[Source repositories need] a search engine like Google Scholar”;
- “[Provide] a thorough, single inter/intranet site where links to all subscribed journal publishers are kept in a straightforward, searchable, concise format”;
- “[Insert a] hyperlink on publisher’s site to relevant source data”.

When asked to specify the options, features or functionality that would enhance their repository searches, as well as considering any functionality considered currently to be missing, responses to the StORe survey¹³ also indicated that researchers have specific concerns not only about data standards and the shortcomings and successes of existing services, but are in considerable need of instruction in respect of what actually is available to them.

A monolithic solution to these issues would be inappropriate and unwelcome. Astronomy, for example, is one discipline to have addressed the researcher/data management partnership in earnest during development of the Virtual Observatory¹⁴ initiative. Notably, one respondent to the StORe survey from astronomy remarked

- “Very happy with what we have in astronomy via arXiv/astro-ph and ADS. Please don’t mess with them for the sake of some aesthetic global model”.

These conditions and observations all provided signal messages to the StORe middleware designers, who in the face of diversity were bent on the production of a generic solution.

3. The pilot demonstrator – an assessment

<u>Work package reference</u>	<u>Documents produced</u>
3: Business Analysis	Discipline Variations, UKDA, November 2006 Business Analysis, UKDA, December 2006
4: Pilot Linkage Build	StORe demonstrator: working specification, UKDA, January 2007 System Walkthru, UKDA, March 2007
5: Evaluation	Summative Evaluation Plan, NCeSS, March 2007 Phase One Summative Evaluation Final Report, NCeSS, August 2007

The task of evaluating the StORe pilot demonstrator was awarded to the National Centre for eSocial Science (NCeSS). This summative evaluation, scheduled to take place between March and May 2007, was planned as an assessment of the technical structure, functionality, design and quality of the demonstrator system, and the appropriateness of its ‘common model’ approach. It would also produce recommendations for future development work. Workshops would be used to test the system with representative repository users, with an international consultant engaged to lead the discussion. Two workshops were planned, each lasting approximately three hours and with eight repository users participating on both occasions.

With assistance from the project team member responsible for surveying the Social Sciences, between 400 and 500 researchers were invited to participate, including contacts established during the 2006 survey of researchers. This invitation was twice repeated, with the added inducement of *Amazon* tokens. Yet only two researchers responded, indicating their willingness to take part, and this willingness had evaporated by the time the workshop arrangements were being made.

Due to this considerable difficulty in identifying participants, the evaluation programme was not completed on schedule. Eventually, a single workshop was arranged for 29th June, when two researchers from the Institute for Social and Economic Research (ISER) at the University of Essex and a data librarian from the University of Edinburgh participated, using the Access Grid at Essex and in Manchester.

At a debriefing session in Edinburgh on 1st August, it was agreed with the NCeSS evaluator that the summative evaluation should be extended, with a second workshop to be convened early in 2008, and following completion of the four system enhancements described on page 15 of this report. The report to be delivered in August 2007 would therefore comprise only part one of the full summative evaluation report.

Participation in the second workshop will be reinforced by the identification of a group of users to test the four planned system enhancements, on an ongoing basis. This same group will then continue as a ‘captive’ cohort for the workshop, which will conclude the programme of testing.

Despite the logistical problems encountered, the June workshop produced a mainly positive assessment of the pilot demonstrator. This was particularly encouraging when the three participants had had no previous exposure to the project and the two researchers in particular were unfamiliar with data repositories. Negative observations tended to be limited to cosmetic issues (e.g. the position of some functions on the screen or the need to introduce categories to the metadata list) and some unfamiliarity with the nomenclature used. Such issues are easily addressed. More importantly, feedback from the workshop indicated that the system was perceived as self-explanatory; it was also very probable that the workshop participants would use it should it be made available as a production system.

The most interesting aspect of the observations made concerned the rationale for the middleware and the strength of its user focus. This was also remarked upon following a live demonstration to the Project Management Group (PMG) on 2nd August 2007, where it transpired that both the workshop participants and the PMG believed the main asset of the StORe system to be its fundamental support of collaborative working. Introduced as a tool for enabling researchers to move between source and output repositories, the system was now also perceived as providing a workspace in which researchers could collaborate, either privately or publicly, sharing the output from their research first as data and then as publications stored in ‘collections’ (a term that had

come to be regarded as synonymous with ‘projects’). Such a view suggests that StORe may have a future in engaging researchers in the management of data on a project or research team basis from early in the life of a project. As options for future development work are explored, the potential for the StORe product to function as a data management environment, possibly functioning as a wrapper around institutional or federated repositories, must surely be considered.

4. Options for development and deployment

Work package reference

Documents produced

5: Evaluation

Phase One Summative Evaluation Final Report, NCeSS, August 2007
StORe continuation activities – indicative plan, August 2007

Continuation options and reinvestment strategy

In the project exit/sustainability plan¹⁵ it is made clear that “this project seeks to undertake groundwork preparatory to further infrastructure development”. It was expected to “conclude with recommendations and pilot software” and, if the demonstrator proved successful in the field of social sciences, additional funding would be sought to undertake “further development to provide the functionality in the other domains of the project”. Hence, although funding was finite it was expected that the project would serve as a proving ground with options for further development.

Worthwhile ambitions to build upon and exploit successful projects often founder in the cut and thrust of competitive bidding for new funds, and the value of the initial investment can consequently be lost. Fortunately for StORe, a predicted end of project budget surplus was identified, produced mainly from three particular events:

	<u>Budget</u>	<u>Expend.</u>	<u>Saving</u>
1.] An astronomy surveyor was not appointed (the individual identified retired unexpectedly and the work was undertaken <i>gratis</i> by project partners)	£14,480	Nil	£14,480
2.] Only a part payment was made for the Biochemistry survey, reflecting the failure to deliver a final report	£14,480	£4,827	£9,653
3.] The Project Manager was appointed after the project commenced, at a lower grade than budgeted, and resigned prior to project completion	£83,949	£58,849	£25,100
Sub-totals	£112,909	£63,676	£49,233
Adjustment (outstanding expenditure) and <u>total saving</u>		£3,000	£46,233

A proposal that these unspent funds should be used to improve upon the initial deliverables from the StORe middleware development has been approved by the JISC. Specifically, four continuation activities with an estimated cost of £30,000 had been identified that will build upon work commenced under the original plan. These, and the rationale for pursuing each of them, are

1. *Enable and test user authentication, initially using Athens but with transition to Shibboleth once the JISC federation is operational* (six weeks). Privacy and protection were established as a fundamental principle during the survey of researchers. Testing authentication will take the pilot closer to an acceptable operational mode.
2. *Produce a top level 'idiot's guide' to the structure of the StORe middleware and embed this as a user-friendly navigational feature of the middleware* (six weeks). This development was inspired by comments from the Project Board and by sample users. Again, it prepares the middleware for a broader context than allowed by the pilot.
3. *Investigate options for decoupling the middleware from dependence on an institutional repository platform, including measures for the middleware to hold the bi-directional links* (ten weeks). An artificial environment was created in which to test the pilot demonstrator. Decoupling it from that environment will show (or disprove) its suitability for export across other institutional loci.
4. *Design and implement a second stage to complete the data deposit process, with the creation of functions to govern the aspects of approval that follow verification* (nine weeks). The focus of the pilot phase was to demonstrate a functional link. A full data deposit process, whilst crucial in an operational

environment, was never constructed, but will need to be proved prior to export of the middleware.

With respect to activity 3, a report on the issues arising will be provided, including the results of a dialogue with other data centres concerning the requirements of their own infrastructures and environments. Support to this dialogue will be provided from the Digital Curation Centre, which is already engaged in an active programme of synthesis and community building between the UK data centres.

As with the original project, a schedule and set of work package summaries will be produced to highlight all dependencies, show anticipated deadlines and deliverables, and to identify who will be undertaking the work. A brief risk analysis will also be supplied.

Environmental testing

In June 2007, within a bid to fund the eCrystals Federation project (JISC Repositories and Preservation Programme, 2007 - 2009), the University of Southampton described the second aim of its work package 2 as the formalisation of the relationship between data in federation and subject repositories and scholarly articles. Earlier that month, at the *Digital Deluge* conference in Manchester, the JISC had confirmed its intention to support a joint development by the StORe and CLADDIER¹⁶ projects and in the eCrystals bid it was proposed that the eCrystals Federation would directly support this initiative by acting as a testbed for the StORe middleware and the CLADDIER 'Ping' mechanism. For StORe this would provide a second domain in which to trial pilot middleware previously demonstrated only within Social Sciences; it would also afford an opportunity to investigate the functionality of bi-directional links with a 'publisher's repository', which the StORe survey confirmed was a high priority deliverable amongst scientific researchers.

The eCrystals Federation work package 2 describes its deliverables as the publication of the Federation Application Profile, an evaluation of the effectiveness of the StORe middleware in the chemistry domain, an analysis of the issues and impact of bi-directional links with a publisher repository, and future options for integrating the StORe middleware in the publishing and research data processes. It is anticipated by the proponents of eCrystals that third party services will develop across a repository federation infrastructure, and the eCrystals model is being developed with this in mind. Whilst such service development will be the subject of future proposals, integration with the StORe project middleware and the CLADDIER Ping mechanism will provide an opportunity for early testing of services that link data and publications.

With respect to the joint StORe/CLADDIER development announced at the *Digital Deluge* conference, the JISC held preliminary talks with the management of both project teams in June. A project scoping exercise is planned for 28th September 2007.

5. Afterword: the project management experience

Project StORe was completed on time, with all work packages executed and deliverables served. That said, a number of practical lessons were learned during the two-year term, which should be of interest to the project sponsors and to subsequent projects.

5.1 Timing - project manager, survey team and project schedule

The award of project funding was confirmed by the JISC in June 2005, with the project term given as 1st June 2005 – 31st May 2007, but with the need to recruit a full time project manager it was impossible to mobilise the project close to the prescribed date. Advertising failed to secure a suitable appointee and it was not until a second round of recruitment in the autumn that a project manager was identified and installed. Although the project manager did not commence his duties until mid-November 2005, the JISC was unable to revise the project term to correspond exactly, agreeing only to new start and end dates of 1st September 2005 and 31st August 2007. Consequently, 2.5 project months were lost.

This had a knock-on effect to the survey of researchers, where the recruitment of team members had been postponed until the appointment of a project manager to lead them. Hence the schedule of work was also affected, with the survey eventually being compressed from the twelve months originally planned to a new programme reduced to just six months. This in turn generated its own difficulties, since critical project milestones came to clash with other events in the academic calendar that otherwise might have been avoided. The limited response to the online questionnaire was in part due to the timing that was forced upon it (resulting in a clash with vacation and exam periods), and the timetabling of interviews suffered from other conflicts of commitment to examinations and research. Both contributed to delays in the delivery of the survey reports.

It is strongly recommended that the JISC should recognise the need for a mobilisation phase at the commencement of any major project, and that project start and end dates should be dependent upon the assembly of a project establishment, accepting that this must be achieved within reasonable and clearly enunciated time limits.

5.2 Project manager and project team

The appointed project manager had a wide-ranging background in information management and experience in senior university management rather than a detailed knowledge of digital repositories. This produced an emphasis on the achievement of effective project management. Involvement with the technical details of middleware design was conducted at a high level in an operator/contractor mode. In the context of a project consisting of varying strands of activity this proved to be an effective strategy, not least because of the positive relationship that was developed between project management and the partners awarded devolved responsibility for the delivery of specialist technical outputs. Such an approach may have been less successful if the conditions for a productive synergy between management and technical experts had been less able to balance.

Five of the six members of the survey team were recruited rather than seconded from existing staff. Four of these five recruits were recent or current PhD students with backgrounds in an appropriate discipline (i.e. their appointed survey domain or information science); the fifth had expertise in market research and relevant domain experience. The sixth member of the team was seconded from the library staff of the partner institution with responsibility for biochemistry. Recruitment of a survey team member for astronomy was unsuccessful, and the astronomy interviews were conducted in the UK by a member of the University of Edinburgh Library team; in the USA, the project management representative fulfilled this role.

The survey team overall was highly and appropriately skilled. The five recruits were particularly motivated by their specific domain and professional interests and this was reflected in the quality of their reports. The seconded member of the team, whilst also contributing effectively and enthusiastically to the project, was increasingly diverted from project business by competing demands from her established post. As a result, completion of the biochemistry interviews and

report were deferred on several occasions and the report was left incomplete when this team member took maternity leave.

From the experience of this project, the appointment of dedicated team members seems to have been the most effective means of delivering against objectives. In the context of a finite project term, with no redundancy in the timetable, the comparative absence of distractions or competing contractual responsibilities appears to have paid dividends.

5.3 Partner obligations and relationships

It was a JISC requirement that project partners should enter into a consortium agreement, which agreement was expected to be confirmed within three months from project inception. A template¹⁷ supplied by the JISC was completed and circulated to project partners with an invitation to sign but the responses were mixed. Whilst some partners were happy to sign, others felt themselves obliged to seek the opinion or authority of institutional legal and/or contractual advisers.

Most partners agreed that a simpler approach was preferable if it would make the legal tone of the document less oppressive and, following consultation with JISC's programme manager, a memorandum of understanding was produced as an alternative. Unfortunately, as this document was required by the JISC to include the same range of essential topics as the consortium agreement, some partners remained obliged to seek institutional advice and in two cases were not authorised to commit to what was perceived as a contractual document.

Notwithstanding the document's importance to the JISC as a formal expression of reassurance that a project can be completed, the programme manager agreed to feed back comments and observations from the StORe experience to JISC Legal. Some clauses were, however, to be regarded as non-negotiable, such as the pre-eminence of HEFCE terms and conditions, the explicit binding of all partners to the JISC terms and conditions in the grant letter, and the clear establishment of Intellectual Property Rights.

5.4 Project wiki

A free *jot.com* wiki¹⁸ was initially created as a mechanism for storing and disseminating project information. The advantage of the wiki over a standard Web site is the ease with which communities can contribute content, which was regarded as important to a project with a geographically distributed team. In the case of the *jot.com* product used for the StORe project, writing to the wiki was found to require minimal technical expertise and there was considerable community and online support.

To enable participation in the wiki, the original free personal plan was replaced in 2006 by the purchase of a group plan. However, contributions from project team and Board/Management Group members proved to be limited, the main function of the wiki site emerging as a more traditional global window on project activities and documentation.

5.5 Bristol Online Surveys

The online survey conducted in 2006 made use of the Bristol Online Surveys¹⁹ licence held by the London School of Economics, one of the project partners. Amongst its general features, options for trialling versions of the StORe questionnaire prior to launch proved especially beneficial, with remote members of the survey team able to provide a commentary and contribute to its development.

Whilst use of the Bristol product was generally intuitive and the product was sufficiently flexible to enable a successful survey, it was quickly discovered that prior experience and skill in designing questionnaires would have been advantageous. The science of constructing a formula that will deliver consistently meaningful responses is not learned without considerable trial and error, and it is recommended that where projects rely significantly on questionnaire data they should if possible employ or invest in appropriate expertise.

5.6 Dissemination programme

A coherent dissemination programme is described in the StORe Project Plan (work package 1), with explicit reference made to project personnel seeking opportunities to present the work of the project at conferences and to publish papers in appropriate professional journals. Whilst this aspiration extended to dissemination events/publications within the seven subject domains, the more significant activity proved to be at events convened within the repository/information management sector, albeit these attracted attendance by representatives from the research community.

One potential outcome from presentations or papers is the generation of sufficient interest to support the export of project outputs or for new collaborations. For StORe, this kind of interest was found amongst the eScience development community, resulting in collaborative partnerships in the UK with the eCrystals Federation and the CLADDIER project, both of which are described above, as well as an invitation for the University of Edinburgh to lead a European consortium developing a bid under the EU's eContent Plus programme, which is yet to be explored.

The presentations and/or papers listed below, which represent the more formal dissemination events in which StORe team members participated, have been deposited in the Edinburgh Research Archive.

Conference presentations and posters

Digital repositories: dealing with the digital deluge, JISC, Manchester, 5 - 6 June 2007

PROJECT StORe – making the link

11th European Conference on Research and Advanced Technology for Digital Libraries, Budapest, 16 - 21 September 2007

ASPECTS OF REPOSITORY USE: employing an ecology based approach to report on user requirements in chemistry and the biosciences [WORKSHOP]

EIPub, Vienna, 14 - 15 June 2007

BEYOND PUBLICATION – a passage through Project StORe

JISC Conference, Birmingham, 13 March 2007

OPENING THE RESEARCH DATA LIFECYCLE: changing roles and responsibilities

IASSIST, Montreal, 16 - 18 May 2007

STORE WARS – may the source and its outputs be with you

Libraries Without Walls 7: exploring 'anytime anywhere' delivery of library services, Molyvos, 14 - 18 September 2007

PROJECT StORe: expectations, a solution and some predicted impact from opening up the research data portfolio

9th International Conference on Asian Digital Libraries, Kyoto, 27 - 30 November 2006

LINKING RESEARCH PAPERS AND RESEARCH DATA: possibilities for a generic solution

OAI5, Geneva, 18 - 20 April 2007

DISSEMINATION OR PUBLICATION? Some consequences from smudging the boundaries between research data and research papers

Open Scholarship 2006: New Challenges for Open Access Repositories, Glasgow, 18 - 20 October 2006

STORE [POSTER]

2nd International Digital Curation Conference, Glasgow, 21 - 22 November 2006

ATTITUDES AND ASPIRATIONS IN A DIVERSE WORLD: the Project StORe Perspective on Scientific Repositories

ADDING VALUE TO THE RESEARCH PROCESS: Project StORe [POSTER]

WWW2006 Repositories Workshop, Edinburgh, 22 May 2006

STORE: linking research papers and research data - possibilities for a generic solution

Publications

Attitudes and Aspirations in a Diverse World: The Project StORe Perspective on Scientific Repositories, *The International Journal of Digital Curation* Issue 1, Volume 2, 2007, Graham Pryor

Beyond Publication – a Passage through Project StORe, *Proceedings of the 11th International Conference on Electronic Publishing*, 2007, Graham Pryor

Project StORe, *D-Lib Magazine*, Volume 12, Number 9, September 2006, Graham Pryor

Project StORe: expectations, a solution and some predicted impact from opening up the research data portfolio, *Libraries Without Walls 7: Exploring 'anytime, anywhere' delivery of library services*, 2007, Graham Pryor

Project StORe: making the connections for research, *OCLC Systems & Services*, Volume 23, Number 1, 2007, Graham Pryor

Use of digital repositories by chemistry researchers: results of a survey, *Program*, 2007, Panayiota Polydoratou

Use and linkage of source and output repositories and the expectations of the chemistry research community about their use, *Lecture Notes in Computer Science*, Volume 4312, 2006, Panayiota Polydoratou

Use and linkage of source and output repositories: interviews with chemistry researchers, *Aslib Proceedings*, Volume 59, 2007, Panayiota Polydoratou

6. References

- ¹ www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories.aspx
- ² Of the 64 data types found by the StORe survey to be deposited in *source* repositories the majority were described in generic terms, including images, plots, instrument data and telemetry, supplemented by a limited number of more specific types according to discipline (e.g. gene/protein sequences and radiographic analyses in the biosciences).
- ³ *Output* repositories generally contain published articles or other texts, usually comprising publications at a pre- or post-refereeing stage, working papers and PhD theses. Output repositories are frequently referred to as institutional repositories, since they are commonly developed first as an institutional resource, and frequently on the basis of a university library initiative. It was also deemed correct to refer to publisher repositories as belonging to the genre; hence a bundle of online periodicals would be considered to be hosted within an output repository.
- ⁴ The Open Archives Initiative and the OAI Protocol for Metadata Harvesting, which are explained at www.ukoln.ac.uk/distributed-systems/jisc-ic/arch/faq/oai
- ⁵ The Dublin Core metadata element set is described at www.ukoln.ac.uk/metadata/resources/dc
- ⁶ The Biotechnology and Biological Sciences Research Council and the Medical Research Council
- ⁷ OECD (2007) *Principles and Guidelines for Access to Research Data from Public Funding* (www.oecd.org/dataoecd/9/61/38500813.pdf)
- ⁸ UKOLN (2007) *Dealing with Data: Roles, Rights, Responsibilities and Relationships* (www.jisc.ac.uk/media/documents/programmes/digital_repositories/dealing_with_data_report-final.pdf)
- ⁹ *Ibid*
- ¹⁰ RIN (2006) *Researchers and Discovery Services: Behaviour, Perceptions and Needs*
- ¹¹ Key Perspectives (2005) Open access self-archiving: an author study (www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf)
- ¹² Questions 23a, 26 and 26a of the StORe online questionnaire sought views on features that would improve searching and the identification of 'missing' functionality
- ¹³ Individual responses to this aspect of the questionnaire are listed on pages 59-60 and 62-69 of the cross-discipline report
- ¹⁴ Details of the international virtual observatory collaboration are given at www.ivoa.net/
- ¹⁵ Pryor (2005) *StORe Project Plan*, Section 19
- ¹⁶ www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories/project_claddier.aspx
- ¹⁷ www.jisc.ac.uk/uploaded_documents/Consortium_Template_v2_May_05.doc
- ¹⁸ jiscstore.jot.com/WikiHome/
- ¹⁹ www.survey.bris.ac.uk