Preserving repository content: practical tools for repository managers

Miggie Pickton, Debra Morris, Stephanie Meece, Simon Coles, Steve Hitchcock [1]

Abstract

The stated aim of many repositories is to provide permanent open access to their content. However, relatively few repositories have implemented practical action plans towards permanence. Repository managers often lack time and confidence to tackle the important but scary problem of preservation.

Written by, and aimed at, institutional repository managers, this paper describes how the JISC-funded KeepIt project has been bringing together existing preservation tools and services with appropriate training and advice to enable repository managers to formulate practical and achievable preservation plans.

Three elements of the KeepIt project are described:

1. The initial, exploratory phase in which repository managers and a preservation specialist established the current status of each repository and its preservation objectives;
2. The repository-specific KeepIt preservation training course which covered the organisational and financial framework of repository preservation; metadata; the new preservation tools; and issues of trust between repository, users and services;
3. The application of some of the tools and lessons learned from the training course to four exemplar repositories and the impact that this has made.

The paper concludes by recommending practical steps that all repository managers may take to ensure their repositories are preservation-ready.

Key words: repositories; preservation; digital preservation; tools

Introduction

Few people would disagree that preservation of repository content is important. Indeed, the stated aim of most repositories is to provide permanent open access to the material therein. Why, then, have so few repositories implemented practical action plans for long term preservation of their content?

Several reasons may be posited. First, few of the existing preservation tools and services have addressed the specific needs of repositories; in practical terms they have necessitated action that is additional rather than integral to repository workflow. Second, repository content is typically highly varied and complex, while descriptive metadata and file formats are used inconsistently and deposited by those without knowledge or expertise in managing digital assets. Finally, busy repository managers with little, if any, experience in digital preservation have lacked time and confidence to tackle what is perceived as an important but complex and scary problem.

[1] Miggie Pickton, NECTAR Repository Manager, The University of Northampton
Steve Hitchcock, KeepIt Project Manager, University of Southampton
Simon Coles, eCrystals Repository Manager, University of Southampton
Debra Morris, EdShare Repository Manager, University of Southampton
Stephanie Meece, UAL Research Online Repository Manager, University of the Arts, London

The JISC-funded KeepIt project brought together existing preservation tools and services with appropriate training and advice to enable the participating repository managers to formulate practical and achievable preservation plans.

From the point of view of the repository manager, this paper summarises the activities of the KeepIt project, describes the impact that the project has had on the participating repositories, and suggests steps that other repository managers might take towards preservation readiness.

The KeepIt project

Institutional repositories are host to a range of different materials, including research papers, teaching materials, creative outputs and datasets. The four participating repositories, NECTAR[2], EdShare[3], UAL Research Online[4] and eCrystals[5], between them represent all of these output types. The managers of the four exemplar repositories plus an experienced preservation specialist and a technical developer made up the KeepIt project team.

For the repository managers, the KeepIt project started with project meetings and one to one discussions with the preservation specialist, Steve Hitchcock. Each manager was also invited to submit their preservation objectives to the project blog. The blog[6], entitled 'Diary of a repository preservation project', tracked progress in the project throughout its duration. Steve Hitchcock was the blog's lead contributor, setting the project within the broader context of digital preservation, highlighting relevant developments in the sector and summarising the tools and services encountered in the project course. However, the four repository managers also contributed some descriptive and reflective pieces, describing the use of new tools in their own institutional contexts and demonstrating their increasing understanding of the issues and challenges of preservation in repositories. The blog is a significant output of the project and there are a number of references to it in this paper.

While the repository managers focused on their separate preservation needs, Dave Tarrant, the project developer, and his colleagues were making rapid progress on a brand new set of tools to manage an integrated repository preservation workflow. These EPrints plugin tools are not the subject of this presentation – they have already been described elsewhere (Field *et al.*, 2009; Tarrant *et al.* 2010) – but they are significant in that they uniquely offer repository managers the opportunity to embed preservation activity, including format management, risk assessment and storage, within the day to day life of the repository.

The meetings and objective-setting exercise highlighted the preservation needs of the exemplar repositories and informed the design and development of a training course in repository preservation. It was decided that the course would begin with the organisational and financial framework of repository preservation, incorporate sessions on metadata and the new preservation tools, and conclude with a consideration of issues of trust between repository, users and services (Hitchcock 2009c).

The ambitions were that following the KeepIt training course:

---

[2] NECTAR website: http://nectar.northampton.ac.uk/
[3] EdShare website: http://www.edshare.soton.ac.uk/
[4] UAL Research Online website: http://ualresearchonline.arts.ac.uk/
[5] eCrystals website: http://ecrystals.chem.soton.ac.uk/
[6] *Diary of a repository preservation project*: http://blogs.ecs.soton.ac.uk/keepit/

- participants would be able to design preservation plans that met the individual needs of their repositories and understand which tools could be used to support these plans;
- the exemplar repository managers would be in a position to lead the way in applying the tools that were most appropriate to their current needs;
- it would have been demonstrated to preservation novices that long term management of repository content was not only desirable and possible, but also achievable within a realistic time and cost framework.

Thus, the KeepIt course and project would set the managers of institutional repositories on the path towards preservation, understanding the process, knowing the pitfalls and having the tools to overcome them. If a full preservation service was not achievable in the timescale of the project, at least they would have taken major strides toward preservation readiness.

Starting out: the exemplar repositories and their preservation objectives

As noted above, the four exemplar repositories were chosen because between them they contained a wide variety of material from research, arts, science and learning and teaching. It therefore stood to reason that their needs would be different, as would be their preservation objectives.

The structured meetings between repository managers and the preservation specialist established the current status of each repository; its mission; management and reporting structure; policy; approach to planning; budget; tools, services and support; storage; content profile; future plans and growth projections. From each conversation a picture emerged of the whole repository within its institutional context.
These conversations not only served to highlight areas of need for future preservation related action, they also provided the opportunity for repository managers to reflect on their current position and to share their thoughts regarding preservation.

Table 1 below summarises the key outcomes of these discussions.

Table 1. Exemplar repositories: key characteristics and initial preservation objectives

	Repository key characteristics	Repository preservation objectives[7]
eCrystals	Data repository containing chemical crystallography data; provides a national service; externally funded on a five year grant basis; data volume expected to rise from present 700 records (each often with a large number of attached files).	Explore preservation training and action required for small groups and non-archivists; investigate how preservation can be made easy; develop a non-onerous preservation regime for the repository administrator; develop preservation costings for researchers to include in funding bids.

[7] The collected surveys and contributed objectives can be found under this blog tag: http://blogs.ecs.soton.ac.uk/keepit/tag/exemplar-profiles/

	Repository key characteristics	Repository preservation objectives[7]
EdShare	Learning and teaching resources repository; rapidly growing with around 7000 digital objects; highly varied content types and file formats; one of several institutional repositories.	Define preservation needs of prevalent file types and formats for learning and teaching; explore preservation concerns of content providers; explore institutional concerns and policies re preservation; understand allocation of responsibility for preservation between creator and repository.
NECTAR	Research outputs repository; strong reporting function; limited full text content (<100 files); embedded institutional service; no separate budget line.	Define preservation needs of all file types and formats in the repository; have procedures and tools to support these; create documentation to inform stakeholders; spread preservation training and knowledge among colleagues with related interests.
UAL Research Online	Arts repository, customised as part of Kultur project; recently launched as institutional service; audio-visual and multimedia content; around 300 records typically with multiple items attached.	Provide guides to digital preservation; gain knowledge sufficient to advise researchers and advocate to senior staff; produce costing for digital preservation to support development of a business plan; plan and implement a preservation programme for the repository.

It is clear from the above that common themes included tools (especially to deal with a range of file formats and ideally integrated with repository workflow); costs (for supporting business plans and funding bids) and organisational issues (such as institutional and user concerns, advocacy, training and documentation). Other objectives occurred uniquely in response to each repository's institutional context[8].

The KeepIt course

Comprising five modules, the KeepIt course sought to introduce repository managers to the practices, tools and services they would need if they were to successfully preserve their repository content.

The themes of the five modules were as follows:

Module 1: Organizational issues, audit, selection and appraisal
Module 2: Institutional and lifecycle preservation costs
Module 3: Primer on preservation workflow, formats and characterisation
Module 4: Putting storage, format management and preservation planning in the repository
Module 5: Trust, of the repository and of the tools and services it chooses

Approximately 16 repository managers from 11 institutions attended the KeepIt training course. Although the KeepIt Project partners all have repositories built on the EPrints software, the course attracted participants with a range of repository types built on a range of different software platforms. Only one section of the course – that covering the new EPrints' plugins described above – was platform specific; the remainder of the course was directly relevant to all repositories.

[8] A synthesis of the four repositories' objectives is given by Hitchcock (2009b).

It is not the purpose of this paper to go into detail about the different elements of the course, for that the reader is referred to the KeepIt blog and the project pages for the individual tools, but their main value for the repository is summarised in Table 2 below.

Table 2. Tools introduced on the KeepIt training course

Tool	Function	Value to repository	Example of use for repository
AIDA (Assessing Digital Institutional Assets)[9]	Establish whether an institution has the organizational, technological and resource capability to support its digital assets	Discover institutional support for and constraints on the repository; identify imbalance between capabilities	None so far (although EdShare used a precursor – the e-learning Maturity Model methodology at an early stage)
DAF (Data Asset Framework)[10]	Identify, locate, describe and assess research data assets	Identify potential repository content; powerful advocacy exercise; support research data policy making	NECTAR, University of Northampton[11]
DRAMBORA (Digital Repository Audit Method Based On Risk Assessment)[12]	Risk assessment and reporting	Identify, assess manage and mitigate risk to the repository	London School of Economics[13]; UAL Research Online, University of the Arts London[14]
DROID (Digital Record Object Identification)[15] and JHOVE (JSTOR Harvard Object Validation Environment)[16]	Automatic file format identification and characterisation	Verify and validate repository content	eCrystals, University of Southampton[17]
EPrints preservation toolkit[18]	File format profiling; implementation of preservation plans; cloud storage of repository content	Identify 'at risk' file formats; monitor repository content; support collection management	EdShare, NECTAR and UAL Research Online have all installed the EPrints preservation plugin

[9] AIDA project website: http://aida.jiscinvolve.org/
[10] DAF project website: http://www.data-audit.eu/
[11] See Pickton, M. (2010b)
[12] DRAMBORA on the DCC website: http://www.dcc.ac.uk/resources/tools-and-applications/drambora
[13] See Fay, E. (2010)
[14] See Meece, S. (2010a and 2010b)
[15] DROID on SourceForge website: http://droid.sourceforge.net/
[16] JHOVE - JSTOR/Harvard Object Validation Environment: http://hul.harvard.edu/jhove/index.html
[17] See Hitchcock, S. (2010a)
[18] See Tarrant, D and Brody, T (2010)

Tool	Function	Value to repository	Example of use for repository
InSPECT (Investigating the Significant Properties of Electronic Content Over Time)[19]	Establish significant properties of digital objects	With respect to repository stakeholders and their use of repository items, know which properties of items need (and need not) be preserved	None so far, although this can also be used as a methodology to inform input to Plato (see below)
KRDS (Keeping Repository Data Safe)[20]	Categorise benefits and costs of the repository	Justify the repository service; build a business case	eCrystals[21] and EdShare[22], both University of Southampton
LIFE3 (Life Cycle Information for E-Literature)[23]	Predict preservation costs	Evaluate the real cost of digital preservation; support the business case	None so far; Beta version of model only recently available[24]
OPM (Open Provenance Model)[25]	Digital representation of provenance	Digitally record provenance of repository items	None so far
Plato[26]	Preservation planning tool	Formally define preservation requirements for repository objects and identify action needed to preserve them	None so far but the EPrints preservation plugin now gives repository managers the ability to import preservation plans[27]
PREMIS[28]	Data dictionary for preservation metadata	Maintain the appropriate metadata for preservation	Florida Digital Archive[29]
PRONOM[30]	Registry of file formats (PRONOM is the registry which informs DROID)	Assess inherent properties of file formats; select lower risk file formats for long term preservation	eCrystals, University of Southampton[31]
TRAC (Trusted Repository Audit and Certification)[32]	Checklist of criteria for assessing trust in repository	Identify repository strengths and weaknesses	eCrystals, University of Southampton[33]

[19] InSPECT project website: http://www.significantproperties.org.uk/
[20] KRDS2 project website: http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx
[21] See Beagrie, N., Lavoie, B. and Woollard, M. (2010)
[22] Morris, D. (2010a)
[23] LIFE3 project website: http://www.life.ac.uk/3/
[24] See [Anon] (2010) *LIFE3 model beta now available for evaluation*.
[25] See Moreau et al. (2010)
[26] Plato preservation tool: http://www.ifs.tuwien.ac.at/dp/plato/intro.html
[27] See Hitchcock, S. (2010b)
[28] PREMIS: Preservation metadata maintenance activity: http://www.loc.gov/standards/premis/
[29] See Donaldson,D.R. and Conway,P. (2010)
[30] The National Archives: the technical registry PRONOM: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx
[31] See Hitchcock, S. (2010a)
[32] Trustworthy Repositories on the DCC website: http://www.dcc.ac.uk/resources/tools-and-applications/trustworthy-repositories
[33] See Hitchcock, S. (2010c)

Some of the tools and services outlined in the course, for example, DAF, AIDA, LIFE3, and the Plato preservation planning tool from Planets, were pre-existing tools designed for more general use in digital preservation, but their application was focussed here for the first time on the special needs of repositories. These tools are less likely to have been used in repositories already but as a result of the KeepIt course might now be taken up by repository managers. Other tools, such as DRAMBORA from the Digital Curation Centre, were developed specifically for repositories.

Applying the tools introduced in the course: case studies

As a result of the course, each of the exemplar managers applied at least one of the tools to their own repositories.

As an externally funded service, the eCrystals repository manager was particularly interested in the costs and benefits of preserving crystallographic data in a repository. This information is essential for making a business case to a prospective funder. Having been a member of the original **KRDS2** project, the eCrystals team knew that whilst it is relatively easy to set up a new repository, it is in populating it with older data that the costs really mount up. In crystallography, as in most scientific areas, the technologies for data creation and storage have changed rapidly over the last 40 years. Migrating raw data from one format to another may result in data loss, but recreating it from scratch is significantly more expensive. The answer, Coles found, is that "the best possible moment to begin preservation is at the time the experiment is performed and data is generated" (Coles 2010b). Undoubtedly this is a message that must be conveyed to data creators in a way that is meaningful to them and likely to result in positive action - such as storing the data in a repository and applying appropriate comprehensive metadata.
Additionally, the eCrystals team added their two main file formats, CIF (Crystallographic Information File) and CML (Chemical Markup Language) to a local copy of the **DROID** format identification tool. (The official PRONOM registry, which informs DROID, is curated by the National Archives and submissions are carefully controlled[34].) By doing this these file types could be automatically verified and validated, in turn enabling the EPrints preservation plugin to assess the risk to these files and if necessary prompt the creation of a preservation plan.

The **EPrints preservation plugin** was used to particularly good effect in describing the different file formats in EdShare, the learning materials repository. With 65 different file formats already in EdShare, and every chance of this number increasing, there are significant preservation challenges to be overcome. In some respects, this analysis raises more questions than it answers, for example:

- What are the institutional and individual's expectations of the repository?
- Must all resources be available to users at all times?
- What are the storage and bandwidth costs of delivering these resources to end users?
- What risk levels are associated with the different file types?
- What could or should the service offer?
- Is EdShare typical of other educational resources repositories?

The EdShare team have made a commitment already to repeating this analysis in order to monitor the situation over time. They will also continue to seek answers to these important questions (Morris 2010b).

[34] See The National Archives: the technical registry PRONOM. Online submission: http://www.nationalarchives.gov.uk/PRONOM/submitinfo.htm

University of the Arts London is in a similar position. As a specialist arts and design institution, its repository, UAL Research Online, needs to be able to hold, manage and showcase outputs from largely practice-based research. Text documents are in the minority; the repository includes, for example, digital records of exhibitions, paintings, textile designs, events, stage designs, films, costume design, sound art, industrial designs, photography, and sculpture. As would be expected for a specialist arts and design repository, the most common file types held by UAL Research Online were .jpegs (2556 files, as of 10 September 2010). Fifty different file formats were identified by DROID, with many of these occurring in only small numbers. The fourth largest category, however, consisted of unknown or unclassified file types – files that DROID was unable to recognise. This was intriguing, as most of the 'unknown' file types had recognizable and very common extensions, such as .mov or .swf. It has not yet been determined why these files were unidentifiable; is it a problem with the signatures on the files themselves, or has this case study provided useful feedback for the development of the DROID database?

With a less complicated file format profile, and indeed far less full content in the repository, the NECTAR team needed to know whether current content was representative of the university's research output as a whole. A **DAF** project was therefore undertaken at Northampton (Pickton 2010a, 2010b; Alexogiannopoulos 2010). Taking advantage of two graduate interns on four week work placements, the project was conducted over a much shorter period than most previous DAF implementations, but nevertheless generated much useful information. For example, it appeared that in addition to the Microsoft .doc and .docx file types and the Adobe .pdf files now predominant in the repository, researchers also commonly used Microsoft spreadsheets (.xls and .xlsx) and .jpeg for images. There was much greater variation in the file types used for databases, audio and video files – these will require a different approach to preservation planning.
Regarding the storage of research data (not currently a service offered by the repository) researchers exhibited different needs and behaviours throughout the research lifecycle, for example storing files on laptops, shared servers and memory sticks at different stages of a research project. Despite recent moves in the sector toward openness, most researchers were reluctant to share their data in a fully open way.
On a more positive front, one outcome of the DAF project at Northampton has been a rise in awareness of issues surrounding research data management and digital preservation. A new research data policy is being developed which almost certainly will inform future repository policy.

The relationship between DAF and the final tool used by the exemplars, DRAMBORA, is complementary. Both have been developed by the Digital Curation Centre (DCC), but whereas DAF focuses on the researcher, the data and the creation stage of the digital lifecycle, DRAMBORA looks at the repository, the process and the preservation phase of the lifecycle (Donnelly 2010).

Two repositories, at the University of the Arts London (UAL) and the London School of Economics (LSE), used DRAMBORA. Both took a lightweight approach with the tool.

At the LSE the team identified ten risks from the DRAMBORA toolkit which were representative of concerns in different organizational, technical and other locally relevant areas. The purpose was to demonstrate the general state of LSE's digital collections and highlight areas where effort needed to be focused. The thorough and transparent nature of

the audit and the provenance of the tool were noted as strengths of the work, giving it an authority that was convincing to senior management. As a result of the work the team were able to identify risks to the current collections and propose appropriate solutions (Fay 2010).

The UAL repository manager had also hoped to use DRAMBORA to identify the activities and assets of the repository and to identify, assess and calculate the associated risks. It was initially felt that DRAMBORA was appropriate for UAL because it is a self-assessment exercise which can be applied to repositories in infancy, it is appropriate in scale, and it is designed for repositories rather than all the digital assets of an organisation. It was hoped that the outcomes of the DRAMBORA project would enable repository staff to define appropriate risk management measures for the repository.
Unfortunately it transpired that UAL's small repository team found it difficult to complete the full process in and around their daily management activities (the guidance provided for DRAMBORA self-auditors suggests that four to five days of 6 hours each would be required to carry out the full self-assessment). Moreover, some elements of the tool required information beyond the scope of usual repository activity. For example, the first sections of the tool required the auditor to have an in-depth understanding of both high level university policies and institutional IT procedures; this information may more easily have been provided by staff senior to the repository manager. A team based approach, such as that used by LSE, may have been more successful.

Impact of the project: Meeting preservation objectives

Having completed the course and applied the tools, the repository managers revisited their objectives.

The **eCrystals** repository had three major goals in fulfilling its preservation objectives. The first was a short-term task and was to directly assist the management of preservation tasks of a research data repository by a research group through implementation of microservices. Thus the primary file types in the repository (CIF and CML) were identified to the DROID service, facilitating verification and validation as described above. Building these operations into the Eprints software brought preservation action right into the repository. The second, longer-term goal, was to understand and develop the relationship between a research data repository and the host institution or research community in terms of migration of preservation plans. This ranges from the short – medium term in the local repository case to the longer term of the institutional or subject repository. Work towards this objective is ongoing. The final objective was to develop costings for researchers. Initial cost data for the eCrystals repository were published as part of the Keeping Research Data Safe study Beagrie *et al*. 2010); these were subsequently translated into full economic cost terms in the KeepIt project blog (Coles 2010a).

In meeting its preservation objectives, **EdShare** had two priorities.
The first was to identify the most prevalent file types in EdShare and, as a complement to this piece of work, to identify the most prevalent file types in the institutional Virtual Learning Environment (VLE), Blackboard. For the repository, the implementation and subsequent application of the EPrints preservation plugin achieved this goal.
The second priority was to explore and understand the specific institutional concerns of the University of Southampton in the preservation of resources for learning and teaching. This work aligned very well with significant ongoing work to develop the "Southampton Learning Environment" (SLE) – a framework for supporting, delivering and enhancing learning and teaching across the whole University community. Work to develop the SLE is planned to evolve over the coming year (2010-late 2011) especially. Development of a suite of "apps" to

support technology enhanced learning at the University of Southampton will also provide additional context for preservation of TEL resources within and for the institution.

Like EdShare, **NECTAR**'s main objectives were to define the preservation needs of all file types and formats held in NECTAR and to have procedures and tools to support these. As a direct result of the KeepIt training course the NECTAR team used the DAF methodology as described above to undertake an audit of research data at The University of Northampton and upgraded their EPrints software to version 3.2 to accommodate the new tools for identifying file types and assessing preservation risks. A third objective, to ensure that preservation training was offered to the broader repository team, was satisfied by inviting technical, metadata and collection management specialists to appropriate elements of the KeepIt training course. This not only spread the acquired knowledge across a wider pool of people, it also promoted engagement with the preservation agenda.

UAL Research Online shared several objectives with the other exemplar repositories. Like EdShare and NECTAR, UAL Research Online benefitted from the installation and use of the new EPrints plugin – it was immediately discovered that over 200 objects in the repository were unidentifiable. This information is invaluable for future management of the collection. Lack of time and conflicting priorities rendered the completion of a full DRAMBORA risk assessment unfeasible, but the work carried out so far will contribute to a thorough documenting of the repository.

Lessons learned from KeepIt

The average repository manager, typified by those involved in the KeepIt project, has little time to comprehensively address preservation issues. So it is not surprising that the exemplars fell short of fully articulating what it would take to execute a complete preservation plan, of undertaking all of the data management tasks required by such a plan, or even of becoming preservation "experts". These things, whilst desirable, were beyond the scope of the project.

What was accomplished was significant progress toward preservation readiness; achieved through increased knowledge and understanding of the elements of preservation and by the implementation of tools appropriate to each repository's stage of maturity.

To sum up, the project has delivered a number of benefits to participants:

- It has increased repository managers' understanding of the scale and complexity of their repository content;
- It has enabled repository staff to achieve greater engagement with content providers;
- It has provided the opportunity for managers to reflect on their repository's current status and think strategically about its future.

Additionally, the project helped managers to raise awareness (of the repository as well as digital preservation) among repository users, colleagues and managers and provide tangible evidence to contributors and senior managers that repositories indeed take seriously their responsibility to ensure secure preservation of the content entrusted to them.

Above all, it convinced the exemplar managers that it is possible to formulate practical and achievable preservation plans.

Steps to preservation readiness

A key aim of both the KeepIt project and its training course was to demystify repository preservation and render it manageable to those responsible for repositories. A priority was to enable repository managers, now informed about preservation and armed with appropriate tools, to take practical steps toward preserving repository content.

To this end, the following actions are recommended. The order of the steps is not fixed and there will be some overlap.

1. **Know the institutional context**. A preservation plan must meet the needs of the institution and its stakeholders. Find out what potential repository content is being produced and by whom. Consider how it is developed, managed and stored (the DAF tool uses this process for research data management, but the principles may be applied to all forms of repository content). Understand your stakeholders' current ability to support preservation.
2. Develop **preservation policy** appropriate to your institution's and users' needs. Consider the content of your repository: does it *all* need preserving, and for how long? A clear policy will determine the scope of preservation activity and support the repository manager in future decision-making.
3. Make a **business case** for preservation– gain the support of your senior managers and demonstrate that preservation can be achieved at realistic cost. The KRDS and LIFE3 models may be helpful.
4. Identify an appropriate preservation **metadata** schema to describe your institution's types of output. This should be built into the repository software and will form part of the standard workflow.
5. Identify **tools** to support preservation planning and decision-making. Use of an appropriate tool will not only provide evidence for future preservation action but in some cases may also facilitate the action itself. So, for example, the EPrints preservation plugins enable file formats to be identified, characterised and risk-assessed and the Plato tool creates an appropriate preservation plan based on your defined requirements, an evaluation of potential strategies for migration and an analysis of the results of these strategies.
6. Consider **storing** repository content in multiple locations, for example in managed 'cloud' storage services. Repository tools such as the EPrints storage plugin, or services such as DuraCloud, can help.
7. Explain to your depositors the benefits of preservation and how the repository can help. **Promote** the preservation services offered by the repository. Build trust among your user community.

To be successful, preservation activity must be embraced by repository managers and embedded within repository workflows and services. The KeepIt project has demonstrated that this is not only desirable, but also possible.

References

Alexogiannopoulos, E., McKenney, S. and Pickton, M. (2010) *Research Data Management Project: a DAF investigation of research data management practices at The University of*

Northampton. (Northampton: The University of Northampton), http://nectar.northampton.ac.uk/2736/

[Anon.] (2010) "LIFE3 model beta now available for evaluation". LIFE project blog, June 11, 2010, http://www.life.ac.uk/blog/2010/06/11/life3-model-beta-available-for-evaluation/

Beagrie, N., Lavoie, B. and Woollard, M. (2010) *Keeping Research Data Safe 2 – Final Report*, pp.64-68. http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf

Coles, S. (2010a) "Preserving crystallographic data in a digital repository: a cost based analysis". Diary of a preservation project, November 9, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/11/09/preserving-crystallographic-data-in-a-digital-repository-a-costs-based-analysis/

Coles, S. (2010b) "Costs, formats and iPad apps: past-future preservation lessons for a science repository". Diary of a preservation project, November 11, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/11/11/costs-formats-and-ipad-apps-past-future-preservation-lessons-for-a-science-repository/

Donaldson, D.R. and Conway, P. (2010) "Implementing PREMIS: a case study of the Florida Digital Archive", *Library Hi Tech*, Vol. 28, No. 2, pp.273 – 289.

Donnelly, M. (2010) "DRAMBORA: Risk and trust and data management". In: *KeepIt Course 5, 30 March 2010, Northampton,* http://blogs.ecs.soton.ac.uk/keepit/2010/09/29/keepit-course-5-drambora-risk-and-trust-and-data-management/

Fay, E. (2010) "Digital collections risk assessment at LSE: using DRAMBORA". Diary of a preservation project, July 19, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/07/19/digital-collections-risk-assessment-at-lse-using-drambora/

Field, A., Tarrant, D., Rauber, A. and Kulovits, H. (2009) Digital Preservation: Logical and bit-stream preservation using Plato, EPrints and the Cloud. In: *13th European Conference on Digital Libraries (ECDL)*, *27 September 2009, Corfu, Greece*. http://eprints.ecs.soton.ac.uk/17962/

Hitchcock, S. (2009b) "Acting on repository preservation objectives". Diary of a preservation project, December 15, 2009, http://blogs.ecs.soton.ac.uk/keepit/2009/12/15/acting-on-repository-preservation-objectives/

Hitchcock, S. (2009c) "Digital preservation tools for repository managers". Diary of a preservation project, December 18, 2009, http://blogs.ecs.soton.ac.uk/keepit/2009/12/18/digital-preservation-training-for-repository-managers/

Hitchcock, S. (2010a) "Adding chemistry to a file format registry". Diary of a preservation project, September 16, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/09/16/adding-chemistry-to-a-file-format-registry/

Hitchcock, S. (2010b) "KeepIt course 4: putting a preservation plan into EPrints". Diary of a preservation project, September 21, 2010,

http://blogs.ecs.soton.ac.uk/keepit/2010/09/21/keepit-course-4-putting-a-preservation-plan-into-eprints/

Hitchcock, S. (2010c) "KeepIt course 5: tools for assessing trustworthy repositories". Diary of a preservation project, September 29, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/09/29/keepit-course-5-tools-for-assessing-trustworthy-repositories/

Meece, S. (2010a) "Digital preservation, risk management and UAL Research Online". Diary of a preservation project, June 11, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/06/11/digital-preservation-risk-management-and-ual-research-online/

Meece, S. (2010b) "Final thoughts on digital preservation work undertaken at UAL". Diary of a preservation project, October 28, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/10/28/final-thoughts-on-digital-preservation-work-undertaken-at-ual/

Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E. and Van den Bussche, J. (2010) "The Open Provenance Model core specification (v1.1)". *Future Generation Computer System* (In Press), http://eprints.ecs.soton.ac.uk/21449/

Morris, D. (2010a) "Keeping data safe in an educational repository: cost-benefits analysis". Diary of a preservation project, February 22, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/02/22/1006/

Morris, D. (2010b) "Repository file type analysis for educational repositories". Diary of a preservation project, October 25, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/10/25/repository-file-type-analysis/

Pickton, M. (2010a) "NECTAR and the Data Asset Framework – first thoughts". Diary of a preservation project, July 2, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/02/07/nectar-and-the-data-asset-framework-first-thoughts/

Pickton, M. (2010b) "NECTAR and the Data Asset Framework – final thoughts". Diary of a preservation project, September 30, 2010, http://blogs.ecs.soton.ac.uk/keepit/2010/09/30/nectar-and-the-data-asset-framework-%e2%80%93-final-thoughts/

Tarrant, D. and Brody, T. (2010) "Preserv2/EPrints preservation plugins 2010-09-10". EPrints Add-ons, last modified September 19 2010, http://files.eprints.org/581/

Tarrant, D., Hitchcock, S., Carr, L., Kulovits, H. and Rauber, A. (2010) "Connecting preservation planning and Plato with digital repository interfaces". In: *7th International Conference on Preservation of Digital Objects (*iPRES2010*), 19 - 24 September 2010, Vienna, Austria*, http://eprints.ecs.soton.ac.uk/21289/

Acknowledgements

The authors would like to acknowledge the JISC for their support for the KeepIt project. We would also like to thank the many individuals who contributed to the KeepIt training course – both trainers and participants.