Open repositories 2010 - General Session

Title:
Preserving repository content: practical steps for repository managers

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

Extended Abstract:

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?

There could be several reasons. Although a number of preservation tools and services already exist, until now few have addressed the specific needs of repositories; in practical terms they have necessitated action that is additional rather than integral to repository workflow. 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. 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.

The JISC-funded KeepIt project is bringing together existing preservation tools and services with appropriate training and advice on preservation strategy, policy, costs, metadata, storage, format management and trust to enable the participating repository managers to formulate practical and achievable preservation plans.

From the point of view of the repository manager, this presentation 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 to ensure 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, EdShare, UAL

[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

Research Online and eCrystals, are representative of all of these output types[2]. The managers of the four exemplar repositories plus an experienced preservation specialist and a technical developer make 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[3]. These reflections highlighted the preservation needs of the exemplar repositories and informed the design and development of a training course in repository preservation.

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[4] – 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.

With objectives and tools in place, 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[5].

It was planned that following the KeepIt training course, participants would be able to design and implement preservation plans that met the individual needs of their repositories; appropriate tools would be implemented in each of the exemplar repositories; and 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.

---

[2] The eCrystals and EdShare projects have been reported previously at Open Repositories:
Coles, S. and Lyon, L. (2008) The eCrystals Federation. At: *Open Repositories 2008 - 3rd International Conference on Open Repositories, 1-4 April 2008, Southampton, UK*. Available from: http://eprints.soton.ac.uk/50902/ [accessed 25 February 2010].
Carr, L.A. *et al*. (2009) Supporting the 'sharing institution' - Practical steps towards a more open teaching and learning culture. At: *Open repositories 2009 - 4th International Conference on Open Repositories, 18-21 May 2009, Atlanta, USA*. Available from: http://smartech.gatech.edu/dspace/handle/1853/28473 [accessed 25 February 2010].

[3] KeepIt project blog: http://blogs.ecs.soton.ac.uk/keepit/

[4] Field, A., Tarrant, D., Rauber, A. and Kulovits, H. (2009) Digital Preservation: Logical and bit-stream preservation using Plato, EPrints and the Cloud. At: *13th European Conference on Digital Libraries (ECDL), 27 September 2009, Corfu, Greece*. Available from: http://eprints.ecs.soton.ac.uk/17962/ [accessed 25 February 2010].

[5] An outline of the course was blogged here: http://blogs.ecs.soton.ac.uk/keepit/2009/12/18/digital-preservation-training-for-repository-managers/

Impact of the project

A systematic survey of the exemplar repositories

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.

Setting preservation objectives

Repository managers were also asked to set preservation objectives[6]. 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)[7]. Other objectives occurred uniquely in response to each repository's institutional context.

Increasing knowledge about preservation and repositories

Repository managers from both within and outside the project attended the KeepIt training course. This was centred on the tools and services which are available to support repository preservation. Several of these, for example, the Data Asset Framework (DAF)[8], the Assessing Digital Institutional Assets self assessment toolkit (AIDA)[9], LIFE3[10], and the Plato[11] 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. Others, such as DRAMBORA[12] from the Digital Curation Centre, were developed specifically for repositories.
Although the project partners were all using Eprints software for their repositories, the KeepIt course attracted participants with a range of repository types. Only one part of the course – the section covering the new Eprints plugins described above – was software-specific; the remainder of the course was of direct relevance to all repositories.

Meeting preservation objectives

Having completed the course, the repository managers revisited their objectives with a view to applying their new knowledge.

---

[6] The collected surveys and contributed objectives can be found under this blog tag: http://blogs.ecs.soton.ac.uk/keepit/tag/exemplar-profiles/
[7] A synthesis of the four repositories' objectives is given here: http://blogs.ecs.soton.ac.uk/keepit/2009/12/15/acting-on-repository-preservation-objectives/
[8] DAF project website: http://www.data-audit.eu/
[9] AIDA project website: http://aida.jiscinvolve.org/
[10] LIFE3 project website: http://www.life.ac.uk/3/
[11] Plato preservation tool: http://www.ifs.tuwien.ac.at/dp/plato/intro.html
[12] DRAMBORA on the DCC website: http://www.dcc.ac.uk/resources/tools-and-applications/drambora

The **eCrystals** repository has three major goals in fulfilling its preservation objectives. The first is a short-term task and is to directly assist the management of preservation tasks of a research data repository by a research group through implementation of microservices. To this end the primary file types in the repository (CIF and CML) will be identified to the DROID service as part of an investigation into the automatic validation and verification of content. This work meets original objectives 1,2 and 3 (preservation actions for non experts). The second, longer-term goal, is 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 and addresses objectives 3 and 4 (administrating repositories). Finally our original objective 4 was to develop costings for researchers and we have published initial cost data for our repository as part of the Keeping Research Data Safe study[13].

In meeting its preservation objectives, **EdShare** now has two priorities. The first is 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 VLE, Blackboard. Having identified these file types, EdShare staff will work with the technical team to identify the preservation needs for these types. It is likely that EdShare will use the EPrints 3.2 plugin developed by Dave Tarrant for this purpose.
The second priority is to explore and understand the specific institutional concerns of the University of Southampton in the preservation of resources for learning and teaching. This work will align very well with significant ongoing work to develop the "Southampton Learning Environment" – a framework for supporting, delivering and enhancing learning and teaching across the whole University community.

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 is using the DAF methodology to undertake an audit of research data at The University of Northampton and will upgrade their Eprints software to version 3.2 to accommodate the new tools for identifying file types and assessing preservation risks. The DAF project has already raised awareness of preservation issues among the research community and the findings will inform future preservation policy and planning.
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.

[13] Keeping Research Data Safe (Phase 2) report:
http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx

UAL Research online has a different challenge, with most of its content comprising audio, video and other non text objects. UAL are using the DRAMBORA self-assessment tool to identify the activities and assets of the repository and to identify, assess and calculate the associated risks. DRAMBORA is 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. The outcomes of the DRAMBORA project will enable repository staff to define appropriate risk management measures for the repository.

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 LIFE3 model 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.