Preserv Preservation Eprint Services

Laying the Foundations for Repository Preservation Services

Final Report from the PRESERV project

Steve Hitchcock, Jessie Hey, Tim Brody and Leslie Carr

Contact for this report: Steve Hitchcock (sh94r@ecs.soton.ac.uk)
March 7, 2007

Acknowledgements

PRESERV was funded by JISC within the programme Supporting Digital Preservation and Asset Management in Institutions 404 programme, theme 3: Institutional repository infrastructure development http://www.jisc.ac.uk/index.cfm?name=programme_404

PRESERV was based at the University of Southampton and was supported throughout by its partners at the British Library, Oxford University and The National Archives.

To find out more about Preserv go to the project Web site http://preserv.eprints.org/

Table of Contents

Executive Summary

The PRESERV project (2005-2007) investigated long-term preservation for institutional repositories (IRs), by identifying preservation services in conjunction with specialists, such as national libraries and archives, and building support for services into popular repository software, in this case EPrints.

We began by producing a simple model showing how preservation services might support repositories. This model changed quite substantially as a result of the project's findings, towards a more powerful and flexible framework that suggests a range of granular Web-based services and providers (Figure 1, from Hitchcock *et al*. 2007a).

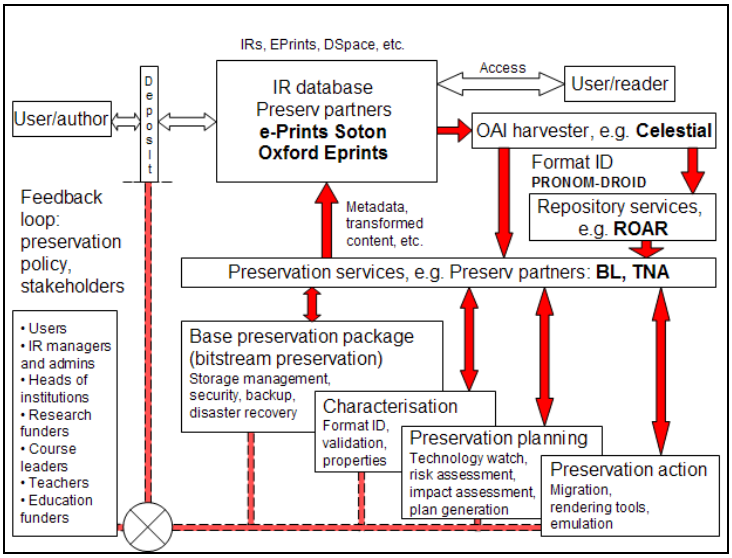

Figure 1. Schematic of latest Preserv service provider model showing granular Web-based preservation services

At the heart of most digital preservation activity is the need for accurate identification of the format of the original source objects. Formats are essentially the signatures of the applications used to create the objects, and because applications change over time to exploit the capabilities of new technology, older versions of digital objects can become unreadable. One approach to this problem is to migrate the original format to a current, readable version. By knowing the formats of all objects in a repository, preservation strategies can be planned and action taken at the appropriate time on those objects that may otherwise be at risk of becoming obsolete.

PRESERV was able to work with The National Archives, which has produced PRONOM-DROID, the pre-eminent tool for file format identification. Instead of linking PRONOM to individual repositories, we linked it to the widely used Registry of Open Access Repositories (ROAR), through an OAI harvesting service. As a result format profiles can be found for over 200 repositories listed in ROAR, what we call the PRONOM-ROAR service (Brody *et al*. 2007).

The lubricant to ease the movement of data between the components of the services model is metadata, notably preservation metadata, which informs, describes and records a range of activities concerned with preserving specific digital objects. PRESERV identified a rich set of preservation metadata, based on the current standard in this area, PREMIS, and where this metadata could be generated in our model. We found that PREMIS appears to provide an excellent basis on which to assess the needs of IRs with respect to preservation metadata, and it is possible to map the PREMIS elements to an extended model incorporating preservation services (Hitchcock *et al*. 2007b).

The most important changes to EPrints software as a result of the project were the addition of a history module to record changes to an object and actions performed on an object, and application programs to package and disseminate data for delivery to an external service using either the Metadata Encoding and Transmission Standard (METS) or the MPEG-21 Part 2: Digital Item Declaration Language (DIDL). One change to the EPrints deposit interface is the option for authors to select a licence indicating rights for allowable use by service providers or users, and others.

Through the results of a survey of repository preservation policy and activity we have a better understanding of what IRs are actually doing to prepare for preservation, and this survey will help service providers to target appropriate services at repositories (Hitchcock *et al*. 2007c).

PRONOM-ROAR changes the outlook for preservation services. By making format profiles openly available and demonstrating Web services as a way to deliver the information, repositories can make more informed decisions about preservation services, and service providers can begin to interact with repositories in more flexible ways. This suggests a range of granular services and providers, starting with example services such as PRONOM-ROAR, which can be tailored to suit the needs of diverse institutions and their repositories, from the largest research university to the smallest teaching college.

PRESERV has identified a powerful and flexible framework in which a wide range of preservation services from many providers can potentially be intermediated to many repositories by other types of repository services. It is proposed to develop and test this framework in the next phase of the project.

Background

"It is important to build the concept of preservation from the outset" (JISC Circular 4/04, note 10). In the digital era, the 'outset' for most new research and educational materials will be the institutional repository (IR).

Since 2002 IRs have seen growth accelerate both in the number of repositories and in the volume of content (Figure 2): between 2005 and 2006 this growth was shown (Suber 2007) to be 25% (repository numbers) and almost 60% (content). In addition, the types of documents deposited in IRs have diversified, in terms of purpose, target audience as well as technical formats. With this growth and diversity comes responsibility to manage the content effectively. According to Lynch (2003) IRs represent "an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution."

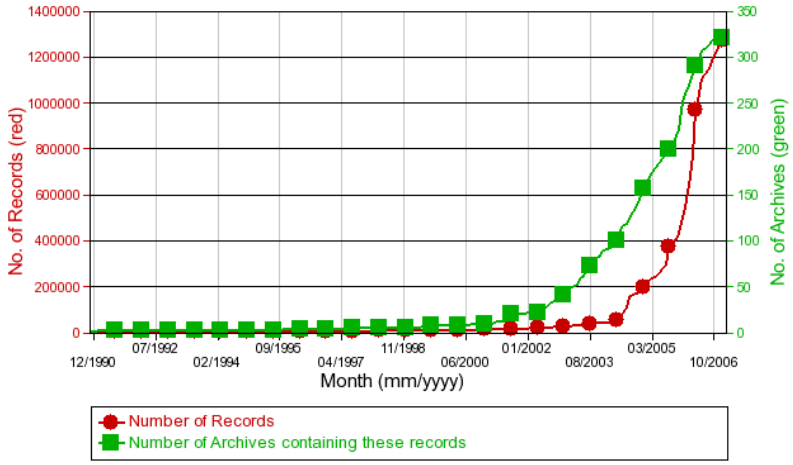

Figure 2. Growth of 'Research Institutional' repositories and contents, generated from the Registry of Open Access Repositories (ROAR) on 5 January 2007

The University of Southampton developed EPrints software, which when it appeared in 2000 was the first software available to build IRs and is now used in over 200 repositories worldwide. EPrints is recognized as having particular strengths for managing digital documents and data of the sort typically produced in the research community, but was possibly perceived as offering less in the way of support for long-term preservation. In comparison with some other IR softwares the differences were, we believed, more of marketing emphasis than substance.

With PRESERV we wanted to investigate the substance of long-term preservation for IRs. We recognised that appearing to emphasise the role of IR software for preservation was potentially misplaced, and omitted organisations with preservation experience and expertise, such as the national libraries and archives. Instead we planned to model and develop preservation services in conjunction with these preservation specialists, and to build support and interfaces for those services into EPrints. This suggested a strong repository-to-repository software relationship with a service provider offering to meet all the preservation requirements of a repository. How we developed this line of thinking was revealed in an interview with the Digital Preservation Coalition (Niven and Hitchcock 2006). What we subsequently identified was a more powerful and flexible framework in which a wide range of preservation services from many providers can potentially be intermediated to many repositories by other types of repository services.

To perform this work the four partners in the project were The British Library (BL), The National Archives (TNA), the Systems and Electronic Resources Service (SERS) at Oxford University, with the School of Electronics and Computer Science at Southampton University as the lead partner.

As part of the JISC 404 digital preservation programme we were also able to work with the Sherpa-DP project, which is similarly investigating preservation services. Our basic models are quite similar: distributed preservation services for IRs. The respective lead partners - Southampton with Preserv, and the Arts and Humanities Data Service with SHERPA DP - see the model from different perspectives, as IR software developer and service provider, respectively. We have been able to learn from another JISC project, Repository Bridge, which successfully demonstrated a preservation service approach for a specific type of

content, electronic theses, showing how these materials could be harvested using OAI and METS from Welsh IRs to a preservation repository at the National Library of Wales.

Aims and Objectives

This section puts the case for each objective set for PRESERV. The section on Outcomes below describes the results.

1 To make our approach to preservation services generalisable to all or most repositories we needed to base it on an accepted standard framework for preservation, in this case the Open Archival Information System (OAIS) reference model

- Implement an ingest service based on the OAIS reference model for institutional archives built using EPrints software.

2 The point of ingest, which in EPrints and other IRs is typically author or administrator-mediated deposit of content, is a primary opportunity to obtain metadata describing the object being deposited. We wanted to identify what metadata would be needed to support preservation services, and where and at what stages in the model and the preservation process those data could be generated.

- Adapt EPrints software to allow the collection and dissemination of preservation-oriented metadata to supplement the current bibliographic information.

3 Preservation processes can embrace many activities, from simple data storage and backup to more complex technical processes. The latter tends to occur because digital data is created using applications that change rapidly in response the increasing power and capability of technology. Applications are characterized by a signature in the form of the format of the object, but as applications change such formats can become obsolete, and unreadable by machines. There are ways of dealing with this problem, migrating older objects to newer formats, for example. To be able to do this requires accurate identification of the original format of a digital object.

- Working with the National Archives, the project will link EPrints to PRONOM software for identification and verification of file formats.

4 To provide test cases for the approaches we would develop in PRESERV we needed to work with real data in real repositories.

- This ingest service will be integrated into the EPrints deposit process for two existing institutional archives, at Southampton and Oxford Universities for evaluation, subject to prior satisfactory testing on pilot archives.

5 If they are to be sustainable preservation services have to be shaped for and marketed to the repository community.

- Consult with stakeholder communities.

6 A preservation service model needs a service provider, a means of demonstrating delivery of the data to be preserved to the provider, and a testbed for the preservation services. Historically, centres of preservation expertise have been based in national libraries and archives, and these centres are now actively developing support for digital preservation, so the BL was a natural fit for this approach.

- The British Library and Southampton University will build and test an exemplar OAI-based preservation service. This service could be used with any OAI-compatible preservation archive to create a software-independent preservation archive.

We will show how and why some of these aims were modified based on the findings of the project. In particular, instead of linking file format identification to repository software it was integrated with a repository registry, making it possible to provide format profiles for many more repositories than anticipated, and revealing a more flexible approach to preservation services.

Methodology

By design, the specific resources available to PRESERV included EPrints software from Southampton University, PRONOM-DROID file format identification software from TNA, test repository environments based on the IRs at Oxford and Southampton universities, and the preservation expertise of the BL.

So the core methodology was to apply these resources to achieve the aims and objectives set out. We would model the proposed approach and attempt to build and test the model in stages, while being prepared to re-evaluate the underlying model according to our findings, and to identify other possible models against which the selected model could be critically compared. The model was to take account of all the aims outlined above, and to provide a basis for related analyses, including mapping preservation metadata, and a survey of repository preservation policy and activity.

We believed at the outset that these resources would be sufficient. The elephant in the room, as it were, was repository services including Celestial (an OAI harvester), and the Registry of Open Access Repositories (ROAR http://roar.eprints.org/) that were being developed separately of the project by a member of the project team. These were to have a large bearing on the eventual outcome of the project.

An omission from the plan, as we were to discover, was the provision of a formal testbed for the preservation services, as commissioned by a service provider. This was not directly funded as part of the project, neither development nor operational costs. Thus a cost-effective facility was not available at the time of the project. A suitable testbed probably would be available now, and a proposal for further work rectifies the earlier costing flaw in this respect.

Implementation

To implement the plan we began by producing a simple model showing how preservation services might support repositories. The evolution of this model is described by Hitchcock *et al*. (2007a), and it changed quite substantially as a result of the project's findings (Figure 1).

At first, however, it had an EPrints repository with PRONOM-DROID in principle applied at the submission interface, and it assumed a storage-based approach to the preservation services, i.e. it assumed a 'black box' in which services were the responsibility of the provider and impinged little on the operation of the originating repository.

Integrating PRONOM-DROID in the EPrints deposit interface was always an uncomfortable proposition. Anything that impacts on the user interface in a repository is sensitive, especially if it might make the deposit process more complex and take more time. It also raised questions about acceptable formats, from a preservation viewpoint, a perennial and still largely unresolved issue, and about when a repository might act on its knowledge of file formats. Should it, for example, provide guidance to authors at the point of submission? If so, that guidance could in principle be to reject a submission and ask the author to reformat.

One answer to the first question about acceptable formats, and which also pointed towards the nature of the preservation services that might ultimately be practical, was suggested by the BL. That answer was to determine the file formats of all existing contents in selected repositories.

There are two ways to build a repository format profile, in each case using PRONOM-DROID to identify the formats:

1. download a complete copy of a test repository
2. obtain an OAI-harvested representation of the repository

The latter is faster and more convenient, although more restricted too. It would only reveal publicly-available content, for example, which would not be sufficient for a repository manager, but it is a reasonable first approximation.

Crucially, the problems with putting format ID in the repository deposit interface, and the idea for profiling content based on harvesting, led to an important new approach. Instead of tying format ID to repositories, through an OAI services approach it could be possible to produce current and updateable format profiles for many repositories. Further, since it is possible to build user services based on OAI harvests, it was possible to envisage an interface enabling the profiles to be viewed by repository managers, preservation service providers, and others. Instead of creating a new interface, a rudimentary interface was added to an existing service, in this case ROAR. As one of the main outputs from the project, PRESERV profiles can be found for over 200 repositories listed in ROAR, what we call the PRONOM-ROAR service (Brody *et al*. 2007).

Before producing PRONOM-ROAR profiles for many repositories, test profiles were produced for two well populated repositories: the University of Southampton repository, and the White Rose repository representing a consortium of the universities of Leeds, Sheffield and York. This helped us to work with TNA to refine the PRONOM registry to include the types of formats found in IRs, since not all were represented, and to evaluate the performance of PRONOM-DROID for completeness, accuracy and ability to differentiate format versions. After some test iterations PRONOM-DROID was ready for use in a live repository service, ROAR.

PRONOM-ROAR changes the outlook for preservation services. A pre-requisite for technical preservation services is file format ID, and it has been shown that, because DROID is open source and ROAR is an open service, such a service can in principle be free. By making repository format profiles openly available and demonstrating Web services as a way to

deliver the information, repositories can make more informed decisions about preservation services, and service providers can begin to interact with repositories in more flexible ways.

Informed by our preservation experts at the BL and TNA, and by PRONOM-ROAR, the black box of preservation services was becoming more transparent. The services that could be made available began to be identified and differentiated, as our model was extended and refined.

The lubricant to ease the movement of data between the components of the model is metadata, notably OAI metadata and preservation metadata. Since EPrints has an OAI-compliant interface, our focus was on preservation metadata, which informs, describes and records a range of activities concerned with preserving specific digital objects. We mapped the current standard for preservation metadata, PREMIS, to the components of our model (Hitchcock *et al*. 2007b): author/IR submitter (via the repository deposit interface); IR software (in this case EPrints); associated tools (in this case file format ID tool PRONOM-DROID); IR policy; and preservation service providers.
.
There are two ways to test the mapping of preservation metadata. First, to apply it to real data that moves about in an implementation based on the model. This was not possible since, as we have already noted, we did not have a suitable testbed for preservation services in this project. An alternative was to test the preservation metadata elements allocated to repositories in the mapping by finding out the extent to which these elements might be used by repositories.

This would also give us the chance to find out how prepared repositories are for preservation. Included in our model was a feedback loop, to enable the model to respond to the needs of stakeholders, in particular to the needs of repository managers, who are at the front-end when it comes to making all aspects of IRs work. There have been surveys of what repository managers and others think about the needs of preservation. Informed by the preservation metadata analysis we produced a more objective survey to find out what repositories are actually doing about preservation, and what policy and provision they are making for preservation.

Preliminary evidence, based on the survey of repositories, shows that most data can be provided by the sources identified, although some elements may need to be adapted or omitted. Beyond the implications for preservation metadata, of more general concern is the finding that no repositories surveyed had a formal preservation policy to guide decisions on the questions raised in the survey, yet *de facto* policies are being applied to specific areas such as file formats, for example.

Through these insights and analyses we began to see the important ways in which EPrints should be adapted to support the evolving model.

Outputs and Results

Starting with the basic IR-preservation services model we were able to analyse and begin to understand the requirements of preservation services. PRESERV has identified a basis for offering flexible preservation services, and has begun to build and test some of the components that will be needed to fulfil these services.

A subset of PREMIS mapped to the services model. We identified a rich set of preservation metadata, based on the current standard in this area, PREMIS, and where this metadata could be generated in our model. We found that PREMIS appears to provide an excellent basis on which to assess the needs of IRs with respect to preservation metadata, and it is possible to map the PREMIS elements to an extended model incorporating preservation services. More implementation and testing are required, especially to validate the allocation of elements to preservation service providers (Hitchcock *et al*. 2007b).

PRONOM-ROAR: a service presenting format profiles of 200+ EPrints and DSpace repositories. This gave us the opportunity to expand file format profiling to over 200 repositories harvested through this service, and to present the results through the widely used interface offered by ROAR (Brody *et al*. 2007). PRONOM-ROAR is also described in a short illustrated guide (http://trac.eprints.org/projects/iar/wiki/Profile).

Preservation features in the latest version of EPrints v3.0 (http://wiki.eprints.org/w/Preservation_Support)**.** Through the preservation metadata analysis we found that hardly any additional data could or should be collected at the point of deposit. The most important changes to EPrints as a result of the project were either internal – the addition of a history module – or on the dissemination side – in the form of Metadata Encoding and Transmission Standard (METS) or the MPEG-21 Part 2: Digital Item Declaration Language (DIDL) plug-ins to package data for delivery to an external service. One change to the deposit interface is the option for authors to select a licence indicating allowable use by users, service providers and others.

- History module. Records changes to an object and actions performed on an object. This will support interaction with service providers, e.g. if a service provider migrates an object to a new format the updated version of the object is returned to the repository after transformation and the history record is updated to show the action taken.
- METS and DIDL plugins. Digital documents, such as HTML Web pages, often comprise a number of digital objects, e.g. text and images. To manage the movement of related objects for purposes such as preservation, they may be packaged. These packages need to be described by metadata, and current standards for these packaging formats include METS and DIDL. In EPrints v3.0 plugins can be written and used to support the transfer of data into and out of a repository in different formats. METS and DIDL plugins are now available.
- Rights declaration. Although we resisted adding format ID to the user interface, one change that made it into the updated interface is a rights declaration. This has wider, legal implications than simply adding something to the EPrints deposit interface, but we learned that some kind of licence statement is necessary for preservation services to act on the content.

Survey of repository preservation policy and activity. These approaches are underpinned by a better understanding of what IRs are actually doing to prepare for preservation, through the results of a survey of preservation policy and activity (Hitchcock *et al*. 2007c). Preservation service providers need to know the scale and shape of the task facing them, and this survey will enable them to understand repositories and help to construct appropriate services. The results were revealing:

- No repositories surveyed had a formal preservation policy

- Preservation policy is being preceded by *de facto* policies on file formats and transformations without provision for acquiring source versions

The first finding can be rectified by good repository policy (not just preservation policy, which will emerge from general considerations). Restricting file formats that can be deposited unwittingly introduces an additional risk factor, because it typically leads to original source data being transformed without documentation, often prior to deposit in the repository. To avoid information necessary to the ongoing process of preservation being lost, restrictions on presentation formats should be accompanied by a requirement for the source version to be deposited.

Extended model of preservation services. If format ID can be provided as a discrete service, is it possible to provide other services as discrete components via Web services, perhaps from multiple service providers? TNA's Seamless Flow programme offered some clues as to the type of discrete preservation services that might be provided. Applying this to PRESERV led to the following structured process for active preservation aimed at repository content, enabling the contributing service components to be identified:

1. Characterisation: identification (as in PRONOM-ROAR), validation, and property extraction
2. Preservation planning: e.g. risk assessment (of generic risks associated with particular formats/representation networks), technology watch (monitoring technology change impacting on risk assessment), impact assessment (impact of risks on specific IR content), preservation plan generation (to mitigate identified impacts, e.g. migration pathways)
3. Preservation action: e.g. migration (including validation of the results) will provide ongoing preservation intervention to ensure continued access or provide on demand preservation action, performing migrations or supplying appropriate rendering tools at the point of user access.

With the addition of a bitstream, or storage-based, preservation package, these services were added to a template for an extended and updated model of preservation services that emerged from a re-evaluation of the original basic model (Hitchcock *et al*. 2007a).

Further work holds promise of realising these services to suit IRs of all types and sizes.

Outcomes

Preservation services are a natural development for institutional repositories, marrying the content expertise of the repository with preservation expertise from the service provider. Despite this preservation services have not yet made a serious impact on IRs, except for particular types of documents, e.g. theses (Hitchcock *et al*. 2007a). To date there are no preservation services aimed at the broad range of documents that might be found in a typical IR. Instead, as a response to the needs of preservation, repositories have been superficially attracted to the preservation features offered by IR software allied to restrictions on file formats that can be deposited.

Preservation services need to be developed to offer an alternative before repositories become compromised by short-term practices. Through PRESERV the role and range of possible

services has become clearer. By fulfilling the aims and objectives set out for the project, if amended in some cases to respond to project findings, some of the required infrastructure required to support services has begun to emerge, including preservation metadata to record both the movement of data between source and services and the actions taken on the data, a format identification process that extends to hundreds of IRs, and support for preservation services built into popular IR software.

In particular we would point to the work with TNA and PRONOM-DROID, the pre-eminent tool in the world for file format identification. Its use in PRESERV was particularly successful, and led to the most important change in all of our objectives. Instead of linking PRONOM to EPrints, and thereby to individual repositories, we linked it to the Registry of Open Access Repositories (ROAR), through an OAI harvesting service.

Working closely with real repositories was important because PRONOM was not designed specifically with IR content in mind, and we were able to identify formats that were not being recognized or that were producing unexpected results. This enabled TNA to modify its registry of formats to provide a more consistent platform for use with a larger number of repositories.

PRONOM-ROAR changes the perspective about the types of preservation services that can be offered and by whom. Instead of a single service providing all conceivable preservation services, it suggests a range of granular services and providers that can be tailored to suit the needs of diverse institutions and their repositories, from the largest research university to the smallest teaching college. So although we didn't build an exemplar preservation service, it was our modeling, and feedback from the BL, that alerted us to the possibilities of a more flexible, interactive model based on distributed Web services, starting with example services such as PRONOM-ROAR.

Our work on modelling preservation services and identifying preservation metadata provided an objective basis on which to ask repository managers what repositories are doing about preservation. We discovered some surprising practices that point to the need for more expert preservation support, and which will inform the construction of those services (Hitchcock *et al*. 2007c).

At the outset an obvious target for PRESERV might have been to produce a marketable and sustainable preservation service offered by an expert institution and aimed at IRs. In retrospect that seems too simplistic. At this stage of birth, growth and development, IRs do not conform to a single model, just as institutions worldwide do not conform to a single model.

Now we have identified a more viable range of preservation services, we have proposed how this work could be extended, working with both the BL and TNA as service providers. In addition, through the recent emergence of international projects such as PLANETS, in which both those partners are involved, we expect to have access to appropriate testbeds in which to construct these exemplar services.

To understand how repositories can respond to these findings, Institutional repository preservation contexts (http://preserv.eprints.org/contexts.shtml) answers some key questions.

Conclusions

PRESERV has identified a powerful and flexible framework in which a wide range of preservation services from many providers can potentially be intermediated to many IRs by other types of repository services.

PRONOM-ROAR is an example of such a service, and provided the motivation and insights to produce the extended services framework. PRONOM-ROAR presents format profiles for over 200 repositories listed in the Registry of Open Access Repositories.

This report has described the method and implementation of the PRESERV investigation, explaining the origins of PRONOM and ROAR and revealing how they came to be combined. It also described some of the components that will help fulfil these services – including a subset of PREMIS mapped to the services model, and preservation features in the latest version of EPrints v3.0 – and reports the results of a survey of repository preservation policy and activity, which will help service providers understand the needs and status of repositories and to construct appropriate services.

Implications

It is time to realise the prospect of preservation services for IRs for the advantages, range and specialist support they could offer compared with limited approaches to repository preservation currently in evidence. An extension of the PRESERV project has been proposed with a view to implementing and testing some of the services identified in the extended model. One of the main limitations of the first phase of the project will be overcome: the project has identified suitable testbeds and multiple providers for the services. Sustainability of services in a market environment is harder to predict.

Recommendations

Preservation planning, rather than preservation action, is the most pressing need for the largest repositories. The good news is this planning need not involve complex technical considerations initially, but instead can evolve from more general considerations – aims, objectives, purpose, audience, etc. – that repository staff will be more familiar with.

- This type of analysis is built into the OAIS approach, for which there is an excellent tutorial (Cornell 2003)
- The OpenDOAR policies tool (http://www.opendoar.org/tools/en/policies.php) will also help with general policy development.
- Use ROAR to try and find a Preserv Profile of formats in your repository.

Repository managers should not panic nor make short-term policy based on preservation considerations alone. We have not detected any immediate threats to the types of content that might be found in IRs at this relatively early stage in their development, other than possibly self-inflicted problems that might be a consequence of file format restrictions.

If a repository restricts file formats that can be deposited, ensure the instruction is accompanied by a request for the source version of the object (this version does not need to be displayed).

Look out for emerging preservation services from specialist providers. There is reason to believe we are moving closer to realising such services. In the UK the BL and TNA are committed to exploring the possibilities of becoming service providers, both within PRESERV, should it be extended, and in other work, The distributed services-based model that has evolved from PRESERV is fully supported by these centres and is consistent with their work elsewhere. In addition, the Arts and Humanities Data Service (AHDS) may offer preservation services to repositories as a result of the SHERPA-DP project.

Be aware of other approaches. For repositories with some local preservation expertise we identified a hierarchy of approaches based on cost (Hitchcock *et al*. 2007a). Other models, including federated network approaches such as LOCKSS, for example, are being investigated elsewhere and might produce viable alternative services.

Reconsider EPrints as preservation-aware repository software. It has new preservation features (http://wiki.eprints.org/w/Preservation_Support) as a result of PRESERV, and through its OAI interface and METS/DIDL plugins for data transfer, which are a feature of the latest EPrints version 3.0 (http://www.eprints.org/software/v3/), is capable of integrating with preservation services.

References

Brody, T., Brown, A., Hitchcock, S., Hey, J. M. N. and Carr, L. (2007) PRONOM-ROAR: adding format profiles to a repository registry to inform preservation services. Preserv project, in preparation

Cornell tutorial (2003) The OAIS Reference Model. In *Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems*, Section 4B, Cornell University, September 2003
http://www.library.cornell.edu/iris/tutorial/dpm/

Hitchcock, Steve, Tim Brody, Jessie M.N. Hey and Leslie Carr (2007a) Digital Preservation Service Provider Models for Institutional Repositories: towards distributed services. *D-Lib Magazine*, May 2007, to be published http://preserv.eprints.org/papers/models/models-paper.html

Hitchcock, Steve, Tim Brody, Jessie M.N. Hey and Leslie Carr (2007b) Preservation Metadata for Institutional Repositories: applying PREMIS. Preserv project, January 2007, submitted for publication http://preserv.eprints.org/papers/presmeta/presmeta-paper.html

Hitchcock, Steve, Tim Brody, Jessie M.N. Hey and Leslie Carr (2007c) Survey of repository preservation policy and activity. Preserv project, January 2007, submitted for publication http://preserv.eprints.org/papers/survey/survey-results.html

Lynch, Clifford A. (2003) Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report*, no. 226, February, 1-7
http://www.arl.org/resources/pubs/br/br226/br226ir.shtml

Niven, Kieron and Hitchcock, Steve (2006) DPC Featured Project, Preserv. Digital Preservation Coalition, 25 July 2006 http://www.dpconline.org/graphics/join/preserv.html

In the course of the project we were influenced and assisted by a wider range of publications than can be represented here, and which can be found in this regularly updated Preserv bibliography
http://preserv.eprints.org/Preserv-bibliography.html

Links

- EPrints software http://www.eprints.org/software/
- Institutional repository preservation contexts http://preserv.eprints.org/contexts.shtml
- JISC 404 programme, Supporting Digital Preservation and Asset Management in Institutions, theme 3: Institutional repository infrastructure development http://www.jisc.ac.uk/index.cfm?name=programme_404
- OpenDOAR policies tool http://www.opendoar.org/tools/en/policies.php
- PLANETS project (http://www.planets-project.eu/
- Preserv project Web site http://preserv.eprints.org/
- Preservation Support in EPrints 3 http://wiki.eprints.org/w/Preservation_Support
- PRONOM-DROID file format identification software from the National Archives http://www.nationalarchives.gov.uk/aboutapps/pronom/tools.htm
- PRONOM-ROAR, An illustrated guide http://trac.eprints.org/projects/iar/wiki/Profile
- Registry of Open Access Repositories http://roar.eprints.org/
- Repository Bridge project http://www.inf.aber.ac.uk/bridge/
- SHERPA-DP project http://www.sherpadp.org.uk/index.html
- The National Archive's Seamless Flow programme http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/default.htm