



UNIVERSITY OF
BATH

Kelly, B., Ashley, K., Guy, M., Pinsent, E., Davies, R. and Hatcher, J. (2008) Preservation of Web Resources: The JISC PoWR Project. In: iPRES 2008: The Fifth International Conference on Preservation of Digital Objects, 2008-09-29 - 2008-09-30.

Link to official URL (if available):

<http://www.bl.uk/ipres2008/programme.html>

Opus: University of Bath Online Publication Store

<http://opus.bath.ac.uk/>

This version is made available in accordance with publisher policies.
Please cite only the published version using the reference above.

See <http://opus.bath.ac.uk/> for usage policies.

Please scroll down to view the document.



University of Bath *Opus*
Online Publications Store

<http://opus.bath.ac.uk/>

COVER PAGE

This is the author's accepted version of a paper made available in accordance with publisher policies.

Author(s): Kelly, B., Ashley, K., Guy, M., Pinsent, E., Davies, R., Hatcher, J.
Title: Preservation of Web Resources: The JISC PoWR Project
Year of publication: 2008

The citation for the published version is:

Kelly, B., Ashley, K., Guy, M., Pinsent, E., Davies, R., Hatcher, J., 2008.
Preservation of Web Resources: The JISC PoWR Project. *In: iPres 2008 Conference*, 29-30th September 2008, British Library, London, UK.

The University of Bath Opus website (<http://opus.bath.ac.uk/>) provides information on usage policies. Please scroll down to view the document.

Preservation Of Web Resources: The JISC PoWR Project

Brian Kelly^{*}, Kevin Ashley⁺, Marieke Guy^{*}, Ed Pinsent⁺, Richard Davis⁺ and Jordon Hatcher^{\$}

^{*} UKOLN, University of Bath, Bath, BA2 7AY

⁺ ULCC, 20 Guilford Street, University of London, WC1N 1DZ

^{\$} Consultant, Opencontentlawyer.com

B.Kelly@ukoln.ac.uk, K.Ashley@ulcc.ac.uk, M.Guy@ukoln.ac.uk, E.Pinsent@ulcc.ac.uk,
R.Davis@ulcc.ac.uk, jordan@opencontentlawyer.com

Abstract

This paper describes the work of the JISC-funded PoWR (Preservation Of Web Resources) project which is developing a handbook on best practices and advice aimed at UK higher and further educational institutions for the preservation of Web sites and Web resources.

The paper summarises the challenges institutions face in preserving Web resources, describes the workshops organized by the project in order to identify the challenges and identify appropriate best practices, and outlines areas in which further work is required.

Background

The preservation of Web resources is a topic that is of interest to many involved in digital curation issues. It presents many interesting technical challenges in terms of capture and access, and organisational and resource-oriented problems, some of which are shared with other aspects of digital preservation and some of which are unique to Web resources. How does one select material? When are we trying to preserve information and when is it the experience, behaviour or appearance that is paramount? How straightforward is it to move Web resources between curatorial environments? Most everyone knows that information persistence on the Web is a fragile thing. And, as Rusbridge has observed [1] even those who care about information persistence don't necessarily do a good job of it on their Web sites. This, despite the fact that good advice about URI persistence has been available for some time [2]. URI persistence is just one small (albeit important) part of the problem that illustrates the wider issues that surround Web preservation in an institutional context.

Not everything on the Web needs to be kept. And there's more than one way to go about keeping it - often it's just the information that needs to survive, and the particular way it is presented on a Web site today is not, of itself, worthy of long-term preservation. Yet there's a lack of knowledge where it's needed about **how** to preserve Web resources, and even when people know how to do it, for some reason it just doesn't happen. That's not a situation the scholarly community is comfortable with, which led to JISC funding the work which is the subject of this paper.

We describe a project funded by the JISC with the aim of producing a series of guidelines on the preservation of Web resources in UK academic institutions. The project, JISC PoWR (Preservation of Web Resources), which is funded from April – September 2008, has established a blog [3] and is running a series of workshops which are helping to gain a better understanding of the challenges institutions face in preserving Web content and support the development of guidelines on best practices.

The paper summarises the work of the project to date, including two workshops which helped to identify challenges and strategies for addressing the preservation of Web resources in a managed Web environment and use of externally-hosted Web 2.0 services.

The project is taking a broad view of what constitutes a Web resource, and hence the remit of the guidelines we will produce. But not everything that is Web-accessible will be covered; for instance, University finance systems will often have a Web interface but are not themselves intrinsically Web resources. But access logs, intranets and externally-hosted content are certainly amongst the types of resource we have been considering, along with the externally-hosted Web 2.0 services which are of growing interest within the sector.

The workshops have endeavoured to bring together institutional stakeholders who might not otherwise encounter each other, such as records managers and Web managers. We are also conscious that it is important to separate decisions about what policy says would be ideal from what is achievable using current resources and technology. We want to bridge the gap between some of the information available about web archiving [16],[17] and their application in a wider organisational context. Where a decision is taken to preserve material, we intend to help institutions make sensible choices between in-house solutions, explicitly-outsourced solutions and what might be described as passive outsourcing: the belief that someone else will do the job for us.

The Preservation Challenges

The Drivers

There are many drivers for undertaking Web site and Web resource preservation within a higher educational institution: institutional policy, legal requirements, and research interests are just a few.

The University is an organisation with business continuity interests that need to be protected. It will have an interest in protecting, managing and preserving certain types of Web content to meet legal requirements and manage its information legislation compliance. The JISC have pointed out that increasingly "*websites may be a unique repository for evidence of institutional activity which is unrecorded elsewhere, and this is often unacknowledged*" [4]. For audit purposes, for example, reference to archived copies of institutional Web sites may be required for the checking of strategic, legal, financial, contractual or scholarly information. If unique records are indeed being created, stored and published on the web, then we'll need to establish their authenticity as records, or as trustworthy and reliable versions of pages.

The University has a responsibility to staff, students, and researchers. Certain services for examinations and assessments are increasingly delivered on the Web; there are static resources accessed through the Web, such as library and learning materials. Students and staff are themselves creators of Web resources, in the form of wikis and blogs; these may sometimes generate content of lasting value. The Web site can be seen as a publication tool, or a dissemination tool; it may be governed by an agreed publication programme. Students will be making career choices, and staff will be making business decisions, based on information they find on the Web - and more importantly, when and where they found it. Does the University have a record of its publication programme? Can it roll back the Web site to a particular point in time to verify what was published two or three years ago? And does it need to be able to roll back the site itself, or the information resource behind the web site?

Research interests are reflected in the increasing number of Web resources that have potential longevity and re-use value, a category that may include scientific research outputs and e-learning objects. Time, money and energy will be wasted if these resources are not preserved, or at the very least protected or managed in some way. There is a heritage dimension and this reflects the University's social responsibility to the academic community; viewed collectively, Web resources will provide interesting insights into the development of Higher and Further Education digital initiatives over the course of the last fifteen years.

Legal Challenges

The preservationist of a digital resource will often find themselves the same legal issues as the original publisher or creator of the resource. This activity, and the others of the preservationist, can carry some legal risks – many of the same risks as the creator of the resources faces in the first place.

Legal issues that can arise when preserving Web resources include:

- Freedom of Information (FOI) legislation, which entitles the public to request recorded information from public authorities, including universities;
- Data Protection Act (DPA) rules governing the use of personal information;
- Intellectual Property Rights (IPRs), particularly copyright;
- Criminal and civil laws that relate to the content of the resource, such as defamation, obscenity, or incitement to racial hatred;
- Contractual obligations such as Terms of Service (ToS) for third party Web sites, particularly in the Web 2.0 space (such as Facebook or Slideshare, mentioned below).

Naturally this list does not exhaust all of the potential legal issues, and each preservation project will have different risks and legal obligations. When examining the potential legal issues on a particular project, it might be useful to break down the issues into the following:

1. **Preservation of a resource because of a legal requirement.** This could be, as mentioned above in a records management context in order to comply with FOI legislation. The "legal requirement" area could be further divided into hard requirements – laws that say something must be retained or preserved – and soft requirements – self-imposed rules to avoid exposure to some legal risk. One example for a soft requirement might be keeping a copy of a Web site's terms and conditions as they evolve in order to prove what terms governed at each exact time.
2. **Legal requirements not to preserve a resource, such as the 5th Data Protection principle:** "Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes" - see [5].
3. **Preservation of content for a non-legal reason but for which legal issues must be addressed.** This could include any number of reasons, such as for cultural heritage.

The notion of **risk management** rather than absolute risk avoidance does however act as an overall umbrella to these three areas. Clearly rules that firmly require information to

be retained or not must be complied with. Concentrating on the possibility of legal liability too much for every area in-between does run another kind of risk – losing the resource.

Engaging With The Communities

The first JISC PoWR workshop took place on 27th June 2008 at the University of London. The day was intended to provide an introduction to the concept of Web preservation and to provide participants with the opportunity to discuss the technological, institutional, legal and resource challenges this presents. The workshop format comprised of a number of presentations and discussion group sessions.

The launch workshop had two primary aims: Firstly to bring together a number of different communities to whom Web resource preservation is of potential importance. This was achieved with an attendance of over 30 people from a wide range of professional groupings, including the Web Management, Records Management and Archives communities. Secondly, to obtain input into the main project goal: the creation of a handbook that specifically addresses digital preservation issues of relevance to the UK HE/FE Web management community. During the day this feedback was provided on the form of suggested content for the handbook, possible delivery scenarios for the handbook and discussion looking beyond the handbook.

The initial presentations and first breakout session explored the challenges that Web resource preservation presents. Consideration was given to the complex nature of the Web: both through its size, transience and reliance on technologies, many of which are external hosted. It was established that Web resource preservation is also hindered by confusion over whose responsibility it is and how decisions on selection should be made. Delegates agreed that one clear requirement for the handbook was the establishment of an effective driver to motivate management buy-in.

The need for fusing of different communities was well demonstrated in the case study presentation given by Alison Wildish and Lizzie Richmond from the University of Bath. Alison (Head of Web Services) and Lizzie (University Archivist, Records Manager and FOI Co-ordinator) described how when asked to give a presentation on their approach to Web resource preservation they had initially felt apprehensive. Although Lizzie could see the value in theory she felt that in practice it was *“too huge a task”*, while Alison admitted that she wasn't really interested and had asked herself *“why is it something I should think about now?”* The task of preparing their presentation, in which they considered the necessary activity of preserving the University prospectus, gave them an understanding of the need for a collaborative approach to the preservation of Web resources.

After lunch a presentation was given on the relevant legal issues Web resource preservation broaches and the suggestion was made that delegates shouldn't panic. A risk assessment approach should be taken and the danger of not preserving should be given a higher priority than legal quandaries.

The second breakout session required delegates to consider possible scenarios related to Web resource preservation. For example one scenario required participants to provide examples of how their organisation's Web site has developed since it was launched. Although there was a lot of 'folk memory' and anecdotal evidence (also known as tacit knowledge) most participants felt they would be unable to reproduce screenshots showing changes to their institution's home page and were forced to rely on third party services, such as the Internet Archive, to provide snapshots of pages on the institutional Web site.

The concluding presentation offered some constructive approaches to protecting an institution's Web site in the short to medium term as part of a records management programme. It was suggested that delegates identify their resources, collaborate with others who have an interest in this area, choose the appropriate approach (or approaches) and accept that the preservation strategy may not, at this stage, include everything. The feedback obtained from attendees during the day will aid in the creation of a blueprint to be given in the project's handbook for the preservation of Web sites and Web resources.

A number of resources were developed for the workshop including three briefing papers on preservation tips, mothballing Web sites and Creative Commons licences. The main presentations were made available via the project blog, with links to audio recording of the talks also provided [6].

Preservation in a Web 1.0 Environment

The Web Managers' Perspective

Sometime in the mid-90s, institutions everywhere seemed to have set up a Web service. At first the service probably contained just a few pages of contact details and institutional overview, although in others cases, departments and individuals may have been able to create their own content sites in sub-sites on a main departmental or institutional service.

Responsibility for managing the Web site may have originated in the Computing Services department, with people skilled in technologies such as HTML, Javascript and CSS. For them term “archiving” would mean creating TAR and ZIP files and painstaking management of sets of daily, weekly, monthly backup tapes. To the 'WebFolk' it

was considerably less likely to mean “keeping a copy of the previous version of the Web site that we can look at again sometime in the future”. This is unfortunate as those early Web sites will have been relatively easy to archive and preserve. By comparison with today's Web resources (which may make use of customisable portals, database-driven services, embedded applications, etc.) collecting a few directories of HTML and JPEG files will have been a trivial task for IT professionals capable of setting up and managing the complexities of Web server software.

Since those early days the Web has grown in sophistication and in complexity. Expectations of design, user interface, content and functionality have grown, for external marketing and publicity services, internal information management on an Intranet and, especially in a Web 2.0 environment, for richer Web-based applications such as, Virtual Learning Environments (VLEs). The Web has now become the platform and interface of choice for virtually every kind of information system.

As we have discovered through our JISC PoWR workshops, Web managers are likely to see their main responsibility as being to their users – keeping online systems useful, usable and up-to-date. That alone requires a lot of running just to stand still. In addition to changing technology and standards, and ever greater demands from creators and consumers of information and publications, there is also an ever-changing regulatory and legislative environment, which may require a complete overhaul of the design of the system.

Therefore it is easy to see why issues that have been identified as key to effective Web preservation – things like persistence, continuity, accessibility, and preservation management – may not be prioritized, or, indeed, even recognised, by members of institutional Web management teams.

Content Management Systems can help with day-to-day management of the Web content. Many even provide version control, though it may be questionable whether such systems could easily recreate a reliable and authentic copy of not only a Web page, but also its environment, functionality, context and embedded external resources. Even if they did, does that commit us to using the same CMS, possibly even the same version, for as long as we want that feature?

What about other systems? Most Web Managers are probably happy to leave responsibilities for management of the content of Web-based institutional repositories, VLEs, discussion boards, etc, to those who requested them. Once again, there are backups, and if any content needs any special attention, each discrete system has a manager whose responsibility that ought to be.

There is just such a huge range of resources on the web it's more than enough to keep a typical Web manager and Web team busy, without them having to consider the nature of records and publications, preservation and archiving as well.

But, as Currall has pointed out [7] that it is simply not a legitimate problem to drop on "the Web guy's" lap, any more than it is one that has an instant technological solution. Deciding what to preserve, and why, is an issue of institutional policy, that needs to be addressed at a senior level across all departments and functions with a Web presence. In universities today, that means everyone.

Armed with a clear brief from policy, Web managers and developers can start thinking about how to capture selected Web objects, and work with the records managers to decide how to store, manage and make them accessible – and what the resource implications of these actions will be

Information Management

The JISC PoWR project proposes that approaches adapted from the information management professions - lifecycle management, records management, archive management - will help with some of the issues raised at the first workshop and discussed in the previous section. We must *manage* resources in order to preserve them (and equally, we must manage them in order to make auditable decisions *not* to preserve them.). An unmanaged resource is difficult, if not impossible, to preserve. Information lifecycle management, if adapted, can help manage Web resources. A records management approach will help to define preservation periods for business records or for legal reasons, even if permanent preservation is not required. Permanent preservation - usually the concern of an archivist - is usually only appropriate for a small subset of resources, for research or cultural purposes.

A records management approach, for example, may be considered suitable when it is known that a Web site contains unique digital records. The Web site itself could be viewed as a record, or - more likely - a potential place where records can be stored or generated. A records manager might ask if people (external and internal) are making business decisions, or decisions about their academic career, based on the information they find on the Web site. Or if transactions, financial or otherwise, are taking place over the Web site and whether the University needs to keep records of these transactions. Are there unique, time-based, evidential records being created this way? If so, how can we capture them?

A Web manager could co-operate with the records manager (and vice versa) to the extent that the site, or parts of it, can start to be included in the University Records Management programme. This may entail a certain amount of *interpretation* as well as co-operation. University policies and procedures, and published records retentions

schedules, will exist; but it is unlikely that they will explicitly refer to Web sites or Web-based resources by name. Where, for example, institutional policies affecting students and student-record keeping are established, we need to find ways of ensuring that they extend their coverage to all appropriate Web resources.

The attraction of bringing a Web site in line with an established retention and disposal programme is that it will work to defined business rules and retention schedules to enable the efficient destruction of materials, and also enable the protection and maintenance of records that need to be kept for business reasons. The additional strength is that the Web site is then managed within a legal and regulatory framework, in line with FOI, DPA, IPR and other information-compliance requirements; and of course the business requirements of the University itself.

The Challenges of Web 2.0

The second JISC PoWR workshop took place on 23rd June 2008 at the University of Aberdeen. This workshop was held as part of UKOLN's annual Institutional Web Management Workshop. The workshop took place after a plenary talk at the event on "*The Tangled Web is but a Fleeting Dream ...but then again...*" given by James Currall [7]. The talk helped to raise the profile of Web preservation for the 180 delegates at the event.

This workshop [8] lasted for 90 minutes. In this short time the discussions and recommendations from the first workshop were described. Participants were then given the opportunity to give their views on a series of scenarios based on use of Web 2.0 technologies including:

- Use of wikis
- Student blogs
- Repository services, such as Slideshare
- Use of Twitter
- Use of Skype
- "Amplified conferences"

The discussions on these particular technologies helped to inform the plans for guidelines on how to address the preservation challenges when making use of Web 2.0 technologies.

Some of the issues that were discussed with regard to these Web 2.0 technologies included:

Wikis: Examples were given of use of externally-hosted wiki services to provide user input, note-taking and user feedback at events. A number of wiki services had been used at a variety of events organized by UKOLN. Typically the wikis were open to anyone for creating and editing the content. This open access policy was taken in order to minimize authentication problems. The

approaches taken to the longer term management of the content was to tighten up the access shortly after the event so that only registered users could edit the content. At a later date only the event organizers could modify the content. In addition the content was migrated from the third party wiki service to a managed environment on the UKOLN Web site.

Blogs: An example of an institutional student blogging service was discussed. Although use of an in-house system might be regarded as allowing the content to be safely managed without the risks associated with use of third party services, there was discussion regarding institutional policies on the management of student data and accounts once the student has left the institution. An example was provided of a student blog which had been migrated from an institutional blogging service to a third party service once the student had left the institution [9]. This example illustrated some of the difficulties in migrating blog content, including bugs in export tools, the limitations of such tools (e.g. only exporting text, and leaving links to embedded content), the loss of blog comments or the difficulties in linking comments with the original blog posts and the difficulties of redirecting the address of the content to new services.

Slideshare: Slideshare is an example of a third party service used for sharing resources – in this case slideshows created by software such as PowerPoint. Although hosting slides on Slideshare has been shown to enhance access to resources [10] there may be concerns over continued access if the Slideshare service is not sustainable over a long period. One approach which has been taken has been to provide a master copy of the slides in an managed environment on the institution's Web site, and to ensure that the title slides and the metadata on the copy on Slideshare provides links to the managed resource.

Twitter: Although many felt that micro-blogging tools such as Twitter should be regarded as personal chat tools with no need to apply institutional preservation policies, it was pointed out that several institutions have already established official Twitter communications channels [11]. In addition UKOLN made use of an official Twitter account to support its IWMW 2008 event, with this technology being evaluated as a possible tool in case of emergencies [12]. There may be a need to take a more managed approach to such technologies used in this fashion. Possible approaches to such management might include the generation of Twitter posts form a centrally-managed service or the harvesting of the RSS feeds from the Twitter service itself. However of more importance than the technical approaches will be to have an understanding of the purposes of the service and the development of preservation policies which reflect those purposes.

Skype: The term ‘Web 2.0’ is now being used to cover a range of technologies including many communications tools. Internet telephony applications such as Skype are now being regarded as Web 2.0 applications, especially when, as is the case with Skype, there are additional applications which integrate with Web services. Is there, then, a need to include such applications when considering how to address preservation of Web resources in a Web 2.0 context? A simple response would be to argue that not only is recording of Skype conversations out-of-scope, the recording of telephone calls without permission may be illegal. However there is a need to consider use of messaging channels which are often provided by such applications. In addition from an institutional perspective it may be desirable to develop preservation policies for digital resources which cover a diversity of technologies and aren’t restricted to Web resources as conventionally understood.

‘Amplified conferences’: Lorcan Dempsey coined the term ‘amplified conference’ to describe events “*are amplifying their effect through a variety of network tools and collateral communications*” [13]. The IWMW 2008 event provided an example of an amplified event, with the provision of a Ning social networking environment, use of Twitter (described previously), a conference back-channel, streaming video of the plenary talks and videos of various informal activities surrounding the event. The variety of technologies which can be used to enhance the effectiveness of an event and increase its impact will provide particular challenges for the preservation of the associated resources. The approaches taken at the IWMW 2008 event have been to (a) document the third party services used, which also supports the event’s approach to risk assessment [13]; (b) migration of appropriate data to managed environments; (c) provision of a diversity of services; (d) use of recommended tags to allow distributed data to be aggregated; (e) recording use of software in cases in which the long term sustainability may be questionable and (f) encouraging use of Creative Commons licence at the event to minimise legal barriers to reuse of the content.

Best Practices for A Web 2.0 Environment

We have described some of the approaches which are being taken to try and address the preservation challenges for an event which is seeking to be innovative in its use of Web 2.0 technologies. But it is acknowledged that the approaches which are being taken by early adopters will not necessarily be easily adopted for use by others. There is a need to document the underlying principles and illustrate how these principles can be implemented.

Why Preserve in a Web 2.0 Environment?

The two main questions which need to be addressed in a Web 2.0 context are the same questions which are relevant

in a Web 2.0 environment: “*Why preserve?*” and “*What are you seeking to preserve?*”. However the diverse ways in which Web 2.0 technologies are being used means that such questions may be more challenging. As we have seen the use of personal and social technologies to support institutional business processes is adding additional complexities to the preservation challenges. And with the diversity of services which are now available and being used for which we cannot guarantee long term sustainability there is a need to be clear as to whether we are seeking to preserve the underlying data, the services used by the institution to fulfill its business processes or the end user experience. There is also the question as to whether it would be acceptable for Web 2.0 services to be lost – a question which may not be understood in, say, a financial context, but may be relevant if services are being evaluated in teaching and learning or research contexts. After all we cannot guarantee that Google will continue to provide a search service, but there are industries which have built services assuming that this will be the case.

Approaches to Preservation in a Web 2.0 Environment

Once the fundamental questions of “why?” and “what?” have been addressed there will be a need to answer the question of ‘how?’. However rather than addressing the specifics of how for particular services some general principles are given below:

Data export: Can the data be exported from the service? Can the rich structure be exported? Can the data be exported in formats which can be imported into other applications or services?

Data import: Can the data be imported into new applications or services? Has the data export / import process been tested? Is any data lost? Do imperfections in the data cause migration difficulties?

Quantifying the costs of migration: What are the predicted costs of migration of the data? How will the costs grow if large-scale data migration is needed?

Content syndication: Can the content be syndicated (using technologies such as RSS or Atom) to allow the content to be made available in other environments?

Sustainability of service: Is the service likely to be sustainable? Are changes to the service likely to be managed gracefully?

Acceptance of risks of loss: Would your organisation be willing to accept the risks of loss of data or a service?

Risks of not using a service: Would your organisation be willing to accept the risks of not using a service (i.e. the missed opportunity costs or the costs of developing or purchasing an alternative service)?

Providing a diversity of content: Is it possible to provide a diversity of content to spread the risks of data loss?

Embedding the learning: The key purpose of a Web 2.0 service may not be the data or the application itself but understanding the underlying processes. The purpose of the service may be complete after the learning has been embedded.

Risk assessment /management: There is a need to develop and share best practices and approaches to risk assessment and risk management.

Raising awareness: There is a need to raise awareness of the importance of preservation strategies.

What Next?

In many respects the challenges of preservation in a Web 2.0 environment have many similarities with preservation in a managed Web 1.0 environment: in both cases there are requirements to clarify why preservation is needed and what aspects of a service need to be preserved. Content managed within the organisation using a Content Management System may appear to be more stable, but we know that Web pages and, indeed, Web site domains, do disappear even from managed institutional environments.

The uncertainties in relying on use of third party services, especially if there are no formal contractual agreements, would appear to make use of Web 2.0 services a risky proposition. But on the other hand since many Web 2.0 services make it easy for content to be created and reused we may find that Web 2.0 services provide a better environment for preserving Web content.

This tension between technologies and approaches which meet immediate business needs and those which best meet long-term policies on information management and retention, is not specific to the web. But the speed with which web services are emerging and evolving make effective decision making more difficult and more urgent than has been the case with other IT developments. Helping institutions define clear, technology-neutral policies and then helping them apply those policies rapidly to emergent systems will be a key success criteria for the guidelines we are developing.

We are also aware that the guidelines may identify a niche for external service provision for the preservation of some web resources. Institutions cannot do everything for themselves; projects such as UKWAC [15], whilst demonstrating the economies of scale that can be achieved in Web archiving, preserve only what their curators select. A number of external service providers exist for web archiving [18] [19] but use of these services by JISC PoWR's target community is vanishingly small. There are a number of possible reasons for this - lack of awareness, cost and an inappropriate service model being amongst them - yet the project has already identified a desire for services broadly like this. Understanding the scale of this requirement for third-party preservation and the ideal

service provision model is outside the scope of what JISC PoWR can achieve today.

There is a need for Web site technologies and management tools to provide better ways of providing long term access to resources, which will include decoupling the address of resources (URIs) from the technologies used to deliver those resources.

But perhaps of even greater importance than technological developments is the need for improved dialogue and shared understanding amongst those involved in developing and implementing policies on Web site preservation.

Life After JISC PoWR

The JISC PoWR project will deliver a handbook on advice and best practices for Web site preservation in an institutional context. But what is the future for Web site preservation after the project's funding ceases? Feedback from the workshops has already encouraged us to view the handbook as a living document, probably hosted on a wiki, rather than as a static publication. This will help to ensure that content remains relevant, although it is no guarantee of continued maintenance.

JISC already ensures that its funded projects are required to document their approaches to the preservation of project resources after the project funded ceases. Recommendations have been made previously by the JISC-funded QA Focus project, and a simple '*Mothballing Web sites toolkit*' was developed [20] to help projects in identifying the policy and technical decisions they would need to make. It might be timely to revisit the development of a more sophisticated toolkit which recognised that projects are likely to make use of Web 2.0 services and ensured that projects had considered the preservation aspects of use of such services.

For institutions, it will be interesting to see whether different approaches to Web resource preservation are equally effective and easy to implement. The project will not last long enough to examine this in depth.

Although the work of the JISC PoWR project has focused on the preservation policies and strategies which institutions should be developing there is also a need to consider the external changes that might be necessary in order to help institutions meet their needs in the most effective manner. The dialogue that the project has enabled between its partners has been fruitful and enlightening for all of us, and it has been rewarding to see similar bridges being built across professional divides as a result of the workshops. We hope that the project's longer-lasting outputs will help to sustain these links and build upon them.

References

- 1 Rusbridge, C. (2008) *RLUK launched... but relaunch flawed?* 21 April 2008, Digital Curation blog, <<http://digitalcuration.blogspot.com/2008/04/rluk-launched-but-relaunch-is-flawed.html>>
- 2 W3C (1998) *Cool URIs Don't Change*, <<http://www.w3.org/Provider/Style/URI>>
- 3 JISC PoWR (2008) JISC PoWR blog, <<http://jiscpowr.jiscinvolve.org/>>
- 4 JISC (2008) *JISC ITT: The Preservation of Web Resources Workshops and Handbook*, <http://www.jisc.ac.uk/fundingopportunities/funding_calls/2008/01/preservationwebresources.aspx>
- 5 JISC Legal Information Service (2007) Data Protection Overview <<http://www.jisclegal.ac.uk/dataprotection/dataprotection.htm>>
- 6 JISC PoWR blog (2008) *Workshop 1 - Resources available*, 30 Jun 2008, <<http://jiscpowr.jiscinvolve.org/2008/06/30/workshop-1-resources-available/>>
- 7 Currall, J. (2008) *The Tangled Web is but a Fleeting Dream ...but then again....*, Presentation at IWMW 2008, <<http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2008/talks/currall/>>
- 8 Guy, M and Kelly, B. (2008) *Approaches To Web Resource Preservation*, IWMW 2008, <<http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2008/sessions/guy/>>
- 9 Kelly, B (2007) *A Backup Copy Of This Blog*, 19 July 2007, <<http://ukwebfocus.wordpress.com/2007/07/19/a-backup-copy-of-this-blog/>>
- 10 Kelly, B (2008) *How Plenary Speakers Are Maximising Their Impact*, 18 June 2008, <<http://ukwebfocus.wordpress.com/2008/06/18/how-plenary-speakers-are-maximising-their-impact/>>
- 11 Kelly, B (2008) *The Open University's Portfolio Of Web 2.0 Services*, 3 July 2008, <<http://ukwebfocus.wordpress.com/2008/07/03/open-university-portfolio-of-web-20-services/>>
- 12 Kelly, B (2008) *Use of Twitter to Support IWMW Events*, 30 July 2008, <<http://ukwebfocus.wordpress.com/2008/07/30/use-of-twitter-to-support-iwmw-events/>>
- 13 Dempsey, L (2007) *The amplified conference*, 25 July 2007, <<http://orweblog.oclc.org/archives/001404.html>>
- 14 UKOLN (2008) *Risk Assessment For The IWMW 2008 Web Site*, <<http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2008/risk-assessment/>>
- 15 Bailey, S and Thompson, D (2006) *UKWAC: building the UK's first public web archive*, Dlib 12(1) <<http://www.dlib.org/dlib/january06/thompson/01thompson.html>>
- 16 Brown, A (2006) *Archiving Websites: A Practical Guide for Information Management Professionals* London: Facet
- 17 Masanes, J (Ed.) (2006) *Web Archiving* Berlin: Springer Verlag
- 18 Hamzo (2008) *Hanzo: The web archiving company* <<http://www.hanzoarchives.com/>>
- 19 Archive-it (2008) *Archive-It home page* <<http://www.archive-it.org/>>
- 20 UKOLN (2005) *Mothballing Web Sites toolkit* <<http://www.ukoln.ac.uk/qa-focus/toolkit/mothballing-01/>>