Archiving The World’s E-Journals
The Keepers Registry As Global Monitor

Peter Burnhill, Françoise Pelle, Pierre Godefroy, Fred Guy, Morag Macgregor and Adam Rusbridge

1. Introduction: The Good News and the Bad

The good news is that so much of the published literature in journals, newspapers and the like can be accessed anytime, anyplace by so many. What is bad news – or at the very least is major cause for concern – is that libraries, whether national or institutional, no longer have custody of that content: their role as stewards for future generations is being undermined.

This is especially true for scholarly and scientific journals, but can also extend to government documents and to every variety of material that libraries have mission to acquire for their patrons. This news is not truly ‘new’ but the stories are still unfolding, and remedial action is required if libraries are to continue as trusted keepers of this key part of our published documentary heritage. There is little doubt that the dynamics in the supply chain have changed: what is now purchased is a strange and uncertain amalgam of ‘rights for access’ and ‘rights to content’. There is a new type of risk to manage when content is online remotely, not on-shelf locally. Different approaches are being considered and acted upon: self-reliance, cooperative action, outsourcing and legal deposit.

The main focus here, and the principal purpose of this paper, is about the measures being taken to ensure that there is continuity of access and assured preservation and integrity for that published content, with special socio-technical challenges faced for serials, in e-journals and the like. This uses the Keepers Registry as a lens with which to highlight the work and achievements of organisations which have stepped forward to act as custodians. The emphasis is placed upon disclosure rather than audit and certification.

Our diverse literature is at risk as cultural product and a store of knowledge, both the content that has been digitized and content that is now born digital and rendered in novel ways. It is vital that we all know who is doing what, and what is not yet done. This is necessarily an international matter as the intellectual product of one nation is important to all others.

The intended outcome of the paper is discussion and agreement on appropriate strategic action, internationally, nationally and locally, to implement archival arrangement for content published.

---

1 Based at EDINA, the UK/JISC national data centre at the University of Edinburgh, Scotland UK, http://edina.ac.uk
2 Based at the ISSN International Centre in Paris, France. The ISSN IC is an intergovernmental institution governed by statutes in a convention between UNESCO and France (the host country). It coordinates the ISSN Network of national centres with the aim of introducing and operating an automated system for the registration of serials, covering the full range of recorded knowledge, through the assignment of International Standard Serial Numbers (ISSNs), http://www.issn.org
3 This is ‘Big Picture’ statement invites qualification related to the ‘Digital Divide’ that occurs within all nations.
electronically in serials. The first priority is that libraries prompt publishers to engage actively with the archiving agencies. The second priority is that those activities are disclosed in the Keepers Registry for all to see.

2. International and National Reports and Activity: Ten Years On

It is almost ten years since the release of the Draft Charter on the Preservation of the Digital Heritage at the UNESCO General Conference 32nd Session in Paris on 19 August 2003, which sought to bring about “a platform for discussions and action on information policies and the safeguarding of recorded knowledge”.

The shift to the digital in the intervening ten-year period has seen a dramatic increase in the number of ISSNs assigned for electronic ‘continuing resources’ by the ISSN Network. This reflects both the growth in e-serials but also the outcome of policy action to assign appropriate identification. In April 2012, the ISSN General Assembly noted that the ISSN Register had a total of about 1.6m entries, of which 97,581 (circa 100,000) ISSNs were for online continuing resources. This includes significant coverage of the major scholarly journals, with one study reporting that 96% of Science journals were online, and as many as 86% of Arts and Humanities were also online. However, the latter statistics may be skewed towards journals that are in the English language and are indexed by Thompson Reuters, omitting significant literature that is published in other languages.

This period has seen a number of stocktaking exercises and reports. What follows is report on investigative activity in the USA and UK. There has been comparable attention in other parts of the world as the significance of this topic is global. Two reports were contemporary with the UNESCO Charter, commissioned by funding bodies in the UK and USA, the Joint Information Systems Committee (JISC) and the Andrew W. Mellon Foundation:

- Archiving E-Journals Consultancy - Final Report by Maggie Jones October 2003
- Archiving Electronic Journals Digital Library Federation 2003

Those reports were wide-ranging and influential, covering topics on licensing as well as preservation. The former report noted in passing that publishers never expected to undertake a preservation role, and they never did with print. Both highlighted the risks associated with digital media and formats (‘digital decay’ such as format obsolescence and bit rot) and with single points of failure: natural disasters (earthquake,}

---

fire and flood) and forms of human folly. The latter include criminal and political action (including hacking whereby unseen changes are made) as well as commercial events associated with the publisher and supply chain, as businesses or product lines end without transfer of legal title, actual content and assured delivery.

3. In Praise of Archiving Organisations

The Draft Charter in 2003 also coincided with the emergence of three types of organizations willing to act as custodians on our collective behalf. Stepping forward with digital preservation programmes are national libraries, cooperative initiatives by research libraries (typically in universities) and third-party initiatives.

The first of these, national libraries, have relied upon legislation for printed publications, requiring publishers to deposit copies of every publication in a given country into at least one library, usually a national library. (In the United Kingdom, for example, legal deposit, as the process is called, has been in existence since 1610.) Typically this enabled both current public access, through a physical visit to a national library, as well as assurance of continuity of access to the printed copy. However, legal deposit of electronic publications is proving to be a complex and difficult challenge, even to this day, in 2012. This is despite the message being put forward by IFLA:12

Legal deposit is critical for the preservation of and access to a nation’s documentary heritage. Publishers and libraries work together to ensure the worldwide success of legal deposit of content, irrespective of format or technology.

The National Library of the Netherlands, Koninklijke Bibliotheek (KB), serves as a pioneering example, with the early emergence of the Depot for the Dutch Electronic Publications (DNEP) having both a national role (for the Netherlands) and with leading publishers based in the Netherlands (Elsevier and Kluwer) also seen to have international significance. Project work began in the late 1990s, including leadership of the EU-funded NEDLIB (Networked European Deposit Libraries) project14 convened by the Standing Committee of the Conference of European National Libraries (CENL). The aim was to define the basic technological conditions for a networked European deposit library. One of the main conclusions, noting the development of the Open Archival Information System OAIS,15 was that archiving should be separated from other access services performed by national libraries, including online search facilities requiring authentication and authorisation.

---

13 In 2002 the KB became the first official digital archive for Elsevier Science e-journals; in 2003 the KB also signed a long-term digital archiving agreement with Kluwer Academic Publishers.
In the USA there was no prospect of national electronic deposit legislation, with concern among policy makers and the libraries for the larger research universities that some other form of action was required. In response and to build upon the work of earlier studies it had funded, the Mellon Foundation provided development funding for initiatives that took two very different approaches to e-journal archiving: the LOCKSS project at Stanford University and the Electronic-Archiving Initiative at JSTOR.

Twelve of the e-journal archiving initiatives that were then current were reviewed in the *Metes and Bounds* report published in 2006. These included Portico (which had emerged from JSTOR), the LOCKSS Network and CLOCKSS (which used the LOCKSS software) as well as digital preservation programmes in the national libraries in Australia and Germany and the province/state-wide action in Ontario (Canada) and Ohio (USA). That report doubted the extent to which the Dutch experience could and should be generalized and would not replace other e-journal archiving agencies to any large extent.

The *Metes and Bounds* report was also influential and although avowedly written from the point of view of academic libraries in North America it contains a useful list of recommendations. Some of these recommendations were directed at publishers and at the archiving agencies, others were for academic libraries/institutions urging that they should become members of, or participate in, at least one e-journal archiving initiative.

Included in that and previous reports was a focus upon establishing the criteria for assessing what is a trusted repository of digital content, with repeated emphasis upon audit and certification. That is not discussed here – but a trail of references can be found at the Digital Curation Centre. There was also a proposal for a registry of archived scholarly publications that indicated which archiving agencies had preserved them, and which publications were still at risk. It is the latter and a focus upon trust through self-disclosure that is discussed below.

### 4. The PEPRS Project and the Keepers Registry

In 2007 JISC commissioned a scoping study on the desirability and feasibility of such a registry, the results of which were positive and indicated support amongst research libraries in the UK. The scoping study recommended that the purpose of the registry should be to enable librarians and policy-makers to discover the archival provision for an e-journal (which archiving scheme has been used, including access/release arrangements) and which e-journals are not (yet) within an archiving scheme. The study considered that it was feasible and recommended that this registry could be built upon SUNCAT, the union catalogue of serials for the UK research and university libraries, considering it might be free at the point of use (‘freely-available’). The evidence collected in the study indicated the importance expressed that the registry should be accurate, up-to-date and comprehensive, and that it was desirable that other

---

16 Archiving Electronic Journals


18 Anne R. Kenney et al. 21-22


20 Sue Sparks, Hugh Look, Adriane Muir and Mark Bide *Scoping study for a registry of electronic journals that indicates where they are archived*. (Rightscom and Loughborough University, 2008). [http://www.jisc.ac.uk/media/documents/programmes/preservation/ejournalregstudy.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/ejournalregstudy.pdf)

21 EDINA is responsible for the delivery of SUNCAT. [http://www.suncat.ac.uk](http://www.suncat.ac.uk)
serials lists and serial union catalogues should be allowed to cross-reference registry entries in order to highlight ‘endangered’ e-journals.

JISC acted on these recommendations. EDINA put together a proposal based upon a partnership with the ISSN International Centre (ISSN-IC), the two organisations having previously worked together in an EU-funded project and since: access to the ISSN Register was regarded as an essential component for the registry. Funding began for a project to “pilot an e-journals preservation registry service” (PEPRS) in August 2008. The aim of the PEPRS project was two-stage: to investigate and then to build and test an online facility that aggregated self-statements by organisations engaging in digital preservation of e-journal content.

A succession of demonstrator systems were created using Ruby on Rails (an open source web framework) in order to cross-match sample metadata supplied by CLOCKSS, e-Depot, Portico and UK LOCKSS Alliance against the authoritative data in the ISSN Register. This enabled some detailed work by ISSN-IC at the serial level and acted as a proof of concept for development of the user interface that searched on journal titles and retrieved information on which, if any, of the agencies included that in their preservation programme, including listings of the actual volumes of journals as held by the agency. Screenshots of the demonstrator were made available on the project website. The development proved successful. An initial reference paper was published, the Abstract Data Model for which is reproduced below.

22 The CASA (Co-operative Archive of Serials and Articles) Project was funded under the Telematics for Libraries' group of projects in the EU Fourth Framework (LB-4058/B-CASA).

23 Peter Burnhill, Françoise Pelle, Pierre Godefroy, Fred Guy, Morag Macgregor, Christine Rees and Adam Rusbridge, “Piloting an e-journals preservation registry service (PEPRS),” *Serials* 22 (2009) 53-59

http://uksg.metapress.com/link.asp?id=350487p5670h0v61
Reports on progress of the project are hosted on the project website\textsuperscript{24} as are various presentations including those made in China, Germany, UK and the USA in order to raise international awareness. The first online release was launched formally in April 2011 at the meeting of the ISSN Governing Board in Paris, as the PEPRS Beta service and generated valuable feedback from across the world. As discussed earlier, we have characterised three types of digital preservation agency: network-level organizations, national libraries, and library co-operatives. Contact had been made with organisations in each category and the following kindly assisted during the project phase:

Two that are network-level organizations:

- CLOCKSS\textsuperscript{25}, a network of archive nodes hosted in libraries around the world managed and overseen by an international Board of libraries and publishers.

---

\textsuperscript{24} “PEPRS” http://edina.ac.uk/projects/peprs/; “The Keepers Registry Blog” http://thekeepersblogs.edina.ac.uk

\textsuperscript{25} “CLOCKSS” http://www.clockss.org/ As disclosure, the University of Edinburgh, of which EDINA is part, is a founder member of CLOCKSS and acts as one of the three Archive nodes in Europe; a further three are in Asia/Pacific, one in Canada and five in the USA: none at present is in Africa, the Arab States or Latin America and the Caribbean.
• Portico:26 operated and overseen by the not-for-profit organization ITHIKA and its Board of Trustees.

Three are managed as part of national libraries:

• e-Depot at KB, Netherlands
• The British Library, UK27
• The National Science Library, Chinese Academy of Sciences

Two are cooperatives formed by libraries, largely in academic institutions:

• The Global LOCKSS Network:28 maintained by Stanford University with funding provided by libraries that join as members of the LOCKSS Alliance.
• HathiTrust:29 a partnership of academic and research institutions own and operate a collaborative digital repository, created to preserve and provide access to millions of volumes digitized from their library collections and other sources; it is administered by Indiana University and the University of Michigan.

The addition of HathiTrust was recognition that the scope of the Keepers Registry had to be wider than the e-journals from publishers and that the digitized content of print journals had to be included. This has also proved insightful for the ISSN Network, both for its decision making on the assignment rules for such digitized serial content and on planning for that assignment. There are many extra challenges of different sort with reporting on archiving of digitized journal content, including problematic volume information as well as lack of identification assignment.

As with all archiving schemes, attention also has to be given to the terms of access. And this is prompt to admit that there is still much to do on establishing agreement on the vocabulary to be used in self-statements about archival policies and practices. This is clearly where the Keepers Registry needs to gain leverage from the work of those who have been developing audit and certification schemes.

5. Enacting the vision

The recommendations for a registry made in the JISC Report in 2003 and the CLIR Report in 2006 became a reality in 2011, at least as a Beta Test Service. That was launched at the ISSN General Assembly in the UNESCO Buildings in Paris in April of that year. With a re-branding and improvement to usability and functionality following feedback, the Keepers Registry service was re-launched, still Beta mode, at the ISSN National Directors meeting in Sarajevo, Bosnia-Herzegovina in October 2011. It is live now at http://thekeepers.org. The screenshot in Figure 2 illustrates the way the Keepers Registry acts as a showcase for the actions being taken by digital preservation organizations, as an online and open record of their archival action on e-journal content.

26 “Portico” http://www.portico.org/
28 “LOCKSS,” http://www.lockss.org/
29 “HathiTrust,” http://www.hathitrust.org/about
Figure 2. The Keepers Registry [screenshot]  
(taken from the Beta release of the Keepers Registry on 29/08/12)

It has search and browse functionality on title, ISSN and publisher. Use of the ISSN-L as the kernel identifier in the system allows search on the ISSN for the print to be entered to establish whether the electronic version of a serial title is being preserved.

The good news is that, by the time of writing (August 2012), seven of the world’s leading archiving agencies report that nearly 18,000 unique Serial Titles are ‘preserved’ - that is one or more actual volumes – and a further 11,000 are ‘in progress’.

Figure 3. Archived content as reported to Keepers Registry [screenshot]  
(taken from the Beta release of the Keepers Registry on 29/08/12)
Figure 3 illustrates the display of the volumes held at the archiving organizations for a given serial. There is some overlap in action across the archiving agencies for this particular serial, which is to be welcomed.

However, there is not good news at the volume level: the extent being preserved for any given Serial varies greatly and is far from complete. Moreover, recall that ISSNs have been issued for 100,000 electronic continuing resources. The job is only part done.

Securing the cooperation of the archiving organisations was and is a priority: these are the organisations that do the work and they have exacting business models that cannot easily absorb extra cost. It was perhaps plain from the outset that the Keepers Registry had to be regarded as an international facility, even though it has been funded by JISC for the UK, as had been predicted in the scoping study:

“It seems to us that in order to gain the co-operation of the archiving organizations based around the world, which would be vital to its utility, the registry/registry-like service would have to be conceived as something which would serve the whole international scholarly community”.

The inclusion of CLOCKSS, e-Depot, LOCKSS and Portico, all of which originate outside the UK, as well as the British Library was both deliberate and essential. This also motivated the early inclusion of the National Science Library of China (NSLC) with which EDINA had previous working contact and exchange of staff. Waiting to join the Keepers Registry are a cooperative organization in Canada, a large-scale university library in the USA and a discipline-specific data organization.

The Keepers Registry seems on course to become a global monitor of international content of significance for each and every country, as well as a showcase for the archiving agencies. Indeed, the Keepers Registry may be that key component in the digital infrastructure that Donald Waters was looking for in 2002 when he wrote:

Our vision is much less clear about the infrastructure needed to enable archives to cooperate and interoperate.  

The good news is that participation in the Keepers Registry prompted discussions amongst the archiving organisations, with calls for greater social media functionality in order that the Registry be function as a ‘safe places network’ enabling exchange between those organisations with archival intent, and also assisting engagement with others who care about the issues involved: not only the titles and extent of volumes actually preserved, but also the variants in technical approach, business model and the terms of availability for triggered content.

In such a conference organised by both UNESCO and IFLA, there should be good advice on hand about how really to be an international facility. There is experience gained via the ISSN Network in dealing with national libraries but we have considered it important to engage with research libraries more generally. There has already been outreach to research libraries through LIBER (for Europe) and ARL (for Canada and the USA). Is there sense in approaching this ‘regionally’, building on the approach made

30 A surprising welcome side-benefit has been contribution to the development of new standards ONIX for Preservation Holdings (ONIX-PH) http://www.editeur.org/127/ONIX-PH/and the Universal Holdings Statement.
31 Sue Sparks, Hugh Look, Adriane Muir and Mark Bide. 32.
with respect to other regions? The approach taken may assist thinking about governance and the long-run business model.

Certainly the use of the ISSN Register within the Keepers Registry is an essential part of this international role as well as what is required for any national role. The connection to the members of the ISSN Network, mostly hosted in national libraries, is also valuable.

The intention to be a global facility has renewed focus upon global usability, and the language in the user interface for the Keepers Registry beyond that of plain English in order to improve the results of automatic translation software, seeking semantic equivalents of Keeper, stewardship and custodian. Feedback on that and offers of help would be welcomed: assistance from UNESCO Volunteers has been mentioned.

6. Forward Look

The literature required and consulted in one country will often have been published by another. This prompts return to the doubts raised in the Metes and Bounds report on the sufficiency of legal deposit and noting following concerns:

1. Differential ability of national libraries to take on digital preservation as part of legal deposit
2. Complexity where a company in one country has an editorial board (and readership) in another, and its servers in a third
3. The resultant haphazard mix of compulsory, voluntary and lack of schemes across the globe
4. Physical access restrictions (having to visit the particular national library).

It is evident that research libraries in the universities have had to act, and not wait upon national libraries. This is true for both for e-journal content that is born digital and for print journals that have been digitized, with obligation to advertise systematically the latter – a role that the Keepers Registry is playing for the content in HathiTrust.

About 20% of the ISSNs issued by members of the ISSN Network for electronic continuing resources have been issued by the Library of Congress with respect to place of publication in the USA. The British Library has assigned about 10% of the total for those published in the UK. Netherlands have each assigned about 4.5%. The archiving organizations being monitored by the Keepers Registry are based in these countries, which in aggregate cover over a third of those digital resources, which have an ISSN. Of course those organizations engage with publishers in other countries. However, there is surely shared priority in university, research and national libraries working together to ensure that publishers in regions across the world engage with the existing archiving organizations. And prompt those organizations, including national libraries that embark upon preservation programmes, to report their activity through the Keepers Registry.

There is another compelling reason why there is no time to wait upon the law to catch up. The focus here has deliberately been upon monitoring archival actions on what is now ‘traditional digital’: scholarly and scientific journals, government documents and the like. However, as the Web becomes the principal arena and medium for scholarly discourse, the problem space has become much richer, as well as more challenging. There is much that is issued on the Web but nowhere else. Scholarly statement and

---

33 Anne R. Kenney et al, Metes and Bounds See pp 21-22
34 The entry in the Keepers Registry for e-Depot: “Licensed content can only be accessed onsite at the KB. Open access and triggered content is freely available via the online e-Depot portal.”
government report contain data and multimedia. This is especially important when considering what is the copy of record.

What is on the Web is now referenced and cited in support of scholarly analysis and statement in e-journals. The Web is dynamic: what is at the end of a given Uniform Resource Locator (URL) can and does change. This topic, termed ‘citation rot’, is being investigated by a team led by Herbert Van de Sompel (Los Alamos National Laboratory) who provides overview in a recorded talk given to the STM Innovations Seminar 2011. It is important to know whether what was referenced has ceased to exist or has been archived somewhere. Using Memento, which enables ‘travel back in time’ by searching the Internet Archive, a recent study reviewed articles in two scholarly repositories in order to establish which cited resources were still current and which were not, and whether the content that existed at the time the citation was made had been archived. “The most dramatic finding is that 45% (66,096) of the URLs that currently exist [in one subject repository] are not archived at all [and 28% of the resources referenced by the articles in the institutional repository had been lost].

This is a challenge, not only for the Keepers Registry.

---


36 “Memento,” [http://www.mementoweb.org/about/](http://www.mementoweb.org/about/)


38 Peter Burnhill, “Tales from The Keepers Registry: Serial Issues About Archiving & The Web,” Serials Review (forthcoming); The University of Edinburgh, including EDINA, have been successful with a funding proposal to work with the Memento Team at Los Alamos National Laboratory to carry out a large-scale investigation with recommendations for archiving cited Web content.