Edinburgh Research Explorer

Who is looking after your e-journals? Telling Tales About The Keepers Registry & Your Digital Shelves

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Who is looking after your e-journals?
Telling Tales About The Keepers Registry & Your Digital Shelves

Peter Burnhill
EDINA, University of Edinburgh, Edinburgh, UK
p.burnhill@ed.ac.uk

Françoise Pelle
ISSN International Centre / Centre International de l'ISSN, Paris, France
pelle@issn.org

Copyright © 2013 by Peter Burnhill and Françoise Pelle. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: http://creativecommons.org/licenses/by/3.0/

Abstract:

The key task for research libraries is to ensure access to the scholarly and cultural record. A significant and growing proportion of that is in digital format and much is found on the Web – and not on the shelves of libraries. This raises important questions about the archival responsibility of libraries and publishers. Our purpose is to report on the current situation of e-journal preservation, on what is being archived and what is at risk of loss. We also indicate strategies that can be considered to meet an international challenge that requires recognition of mutual inter-dependence across the globe. The literature that is consulted and required by researchers in one country will often have been written and published in another country.

The first (and easiest) priority for research libraries is to focus on e-journals and take prompt and strategic action, both to avoid loss in the short term and to establish means to assess progress towards the (achievable) goal of ensuring that there is complete and effective e-preservation plans for all of our e-journal content. This is assisted by The Keepers Registry, http://thekeepers.org, which provides a lens onto the extent of e-journal archiving as the leading archiving agencies report what they have ingested. The sustainability of archiving activity, and the means to monitor that activity, is of major strategic importance.

A related priority is to tackle the variety of ‘serial issues’ that can improve the effectiveness of archiving and monitoring. These include identification (e.g. ISSN and ISSN-L) of all types of continuing resources, particularly journals but also ongoing ‘integrating resources’ such as databases and Web sites; the consistent naming and identification of publishers (e.g. ISNI); and the continuing need for a universal holdings statement for assurance that each and every volume and issue has been successfully archived.

Keywords: Serials, E-journals, Digital Preservation, ISSN, Keepers Registry
Overview

The fundamental task for a library is to ensure ease and continuity of access to information. The special mandate for the research library is to ensure continuity of access to the scholarly record, both for current researchers and students and for future generations. Today the scholarly record is mostly in digital form. This is true for e-journal content that is born digital, and may only exist in digital form, and for the older print journals that are being digitized.

It may be old news that libraries no longer take custody of the content that they buy or that which is available increasingly as ‘open access’. However, despite progress over the past ten years, the evidence presented here strongly suggests that the challenge to provide long-term assured access to e-journals is not being met sufficiently by the research library community.

- Much the same could be said about the cultural record and the obligation on every library to ensure continuity of access for present and future citizens of every country.

- IFLA was an organising partner in Vancouver in 2012 on the occasion of the 20th anniversary of the UNESCO Memory of the World Program.¹ The Vancouver Declaration² recommends “active advocacy role in making digital preservation frameworks and practices a reality, by promoting digital preservation in all appropriate forms” … and urges professional associations “to assist in the development of a cohesive and practical vision of the way forward in addressing the management and preservation of recorded information in all its forms in the digital environment”.

There should be no doubt that research and scholarship are international in scope: literature consulted by researchers in one country will often have been published by researchers in another country; no one is self-sufficient. Digital preservation, with assured access conditions, of e-journals and the like should be a priority for research libraries in each and every country. National libraries have a key role to play with respect to the publishers in their country, reflecting on the archival responsibility of publishers as well as libraries. However, research libraries in the universities and research institutions across the world have had to take the initiative and act without waiting upon national libraries and national legislation.

What was once on-shelf in a library is now on-line somewhere else. We all need to know who is looking after the world’s e-journals for the long-term, how, and with what terms of access? It is now timely for international bodies to generate global awareness of this shared challenge and then to turn this into practical and sufficient action to avoid a real threat of loss.

We will present arguments and evidence, also set out at length in two recently published articles: these are ‘Tales From The Keepers Registry: Serial Issues About Archiving & the

---

Web”³ and ‘Archiving The World’s E-Journals: The Keepers Registry As Global Monitor”⁴. More generally we draw upon work done by many others over the past ten years including that by the archiving agencies (the keepers of digital content) who have worked together with by our colleagues’ over the past five years to deliver the online facility called The Keepers Registry. Developed at the initiative of Jisc in the UK, this Registry was designed and created as an international facility by EDINA (at the University of Edinburgh) and the ISSN International Centre in Paris.

Background Reports

It is 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”⁶.

Two other reports were commissioned in 2003, by the Joint Information Systems Committee (JISC)⁷ and the Andrew W. Mellon Foundation:

- Archiving E-Journals Consultancy - Final Report by Maggie Jones October 2003⁸
- Archiving Electronic Journals Digital Library Federation 2003⁹

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.

These reports and the Metes and Bounds¹⁰ report published in 2006 suggested a registry of archived scholarly publications that would indicate which archiving agencies had preserved them and which publications were still at risk. In 2007 Jisc commissioned a study to investigate the perceived need for such an e-journal preservation registry. It reported and that eventually led to the project funding for EDINA and the ISSN IC to ‘pilot an e-journal registry service’ (the PEPRS project) in 2008. In 2009, Jisc Collections also issued their

---

³ Peter Burnhill, Tales from The Keepers Registry: Serial Issues About Archiving & the Web, Serials Review, Volume 39, Issue 1, March 2013, Pages 3-20, ISSN 0098-7913, 10.1016/j.serrev.2013.02.003. [Also found at http://www.era.lib.ed.ac.uk/handle/1842/6682]
⁵ Project team: Pierre Godefroy, Fred Guy, Morag Macgregor, Christine Rees and Adam Rusbridge.
⁷ “JISC” (now re-badged as Jisc), accessed on 29 August 2012 http://jisc.ac.uk
http://www.jisc.ac.uk/uploaded_documents/ejournalsfinal.pdf
practical guide to e-journal archiving’ in the UK\textsuperscript{11}. Previous reports had included 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 - a trail of references can be found at the Digital Curation Centre\textsuperscript{12}.

\textbf{Our Purpose}

We use The Keepers Registry to report on the current situation of e-journal preservation, on what is being archived and what is at risk of loss. We also indicate strategies that might be considered in order to meet this international challenge, beginning with recognition of mutual inter-dependence across the globe. The Data Model, taken from the article in which we set out our intent to build a Registry\textsuperscript{13}, illustrates the central role of the ISSN and the ISSN Register in the design and operation of The Keepers Registry.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Abstract_Data_Model.png}
\caption{The Abstract Data Model for a Registry of Preservation Action}
\end{figure}

\textbf{In Praise of Our Keepers of Digital Content}

During the project activity for The Keepers Registry we noted that these seem to be three types of organization acting as our digital shelving, as keepers of e-journal content:

i. Organizations that were created with mission to preserve e-journal content, international in intent, operating at ‘web-scale’. The two most notable are the CLOCKSS Archive (http://www.clockss.org/) and Portico (http://www.portico.org/), both of which were participants in the project from the outset.


\textsuperscript{11} http://www.jisc-collections.ac.uk/Documents/practical_guide_to_ejournal_archiving.pdf
\textsuperscript{12} Digital Curation Centre, http://www.dcc.ac.uk/resources/tools-and-applications/trustworthy-repositories (accessed on 30th May 2013)
We anticipate a growing number undertaking e-journal archiving, sometimes supported by legal deposit legislation, sometimes not.

iii. Collaborative library initiatives: UK LOCKSS Alliance, part of the Global LOCKSS Network (http://www.lockss.org/), HathiTrust (http://www.hathitrust.org/about) and, the Scholars Portal from the Ontario Council of University Libraries. We anticipate others, especially with specialist roles such as the Archaeology Data Service (ADS).

Some Evidence

The shift to the digital in the ten-year period since those early reports has seen a dramatic increase in ISSN assigned for electronic ‘continuing resources’ by the ISSN Network. This reflects both the growth in e-serials and the outcome of policy action by the ISSN Network. 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) ISSN were for online continuing resources\(^\text{14}\). This includes significant coverage of the major scholarly journals, with one study\(^\text{15}\) reporting that 96% of Science journals were online, and as many as 86% of Arts and Humanities were also online. The numbers are now probably much higher.

The problem really is international. 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. Canada and the Netherlands have each assigned about 4%. There is then a very long tail.

The Keepers Registry

The recommendations for a registry made in the JISC Report in 2003 and the CLIR Report in 2006 became a reality in 2011, when the Beta Test Service was launched at the ISSN General Assembly in the UNESCO Buildings in Paris. With a re-branding and improvement to usability and functionality following feedback, the Keepers Registry service was re-launched later that year, still Beta mode, at the ISSN National Directors meeting in Sarajevo, Bosnia-Herzegovina, to provide librarians, policy makers and the publishing industry with:

- a global lens onto assured access to the world’s online (serial) literature
- a show-case for archiving agencies, and is developing into
- a spotlight onto digital content that is at risk of loss
- the focal ‘social media’ meeting point for a ‘safe places network’.

Links to more detailed information are provided on the website and in an associated Keepers Blog, http://thekinders.blogs.edina.ac.uk/, including a number of publications and presentations that describe how the Registry came about via the PEPRS project.


http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/e-journals-their-use-value-and-impact
Evidence of Incomplete Archival Activity

At the time of writing (May 2013), nine archiving organizations have signed up the Registry – with another large library waiting to do so. The eight archiving organizations are now reporting into The Keepers Registry affirm that about 20,600 unique titles are being preserved, as indicated in the following screenshot.

Figure 2: Screenshot of The Keepers Registry (as at 31 May 2013)

As yet, there is no categorisation of serials within The Keepers Registry that would enable direct assessment of the extent to which these 20,500 preserved titles are a significant percentage of the 30,000 or so peer-reviewed journals, but clearly this represents only about 20% of the ‘continuing resources’ and ‘integrated resources’ having an ISSN. Another take on such progress is from the point of view of a given research library, with focus on the list of e-journals that it wishes to provide to its students and staff, especially as moves to some variation of ‘digital only’. With their kind permission, the three university libraries have allowed the summary title-level statistics taken from Table 1 in Burnhill (2013).

Figure 3: Using The Keepers Registry To Assess Archival Status (August 2012)

<table>
<thead>
<tr>
<th>University Library</th>
<th>% ‘Preserved’ by 1 or more</th>
<th>% ‘Preserved’ by 3 or more</th>
<th>% Not known as ‘Preserved’</th>
<th>Total having a valid ISSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columbia</td>
<td>26</td>
<td>12</td>
<td>74</td>
<td>58,882</td>
</tr>
<tr>
<td>Cornell</td>
<td>28</td>
<td>13</td>
<td>72</td>
<td>54,698</td>
</tr>
<tr>
<td>Duke</td>
<td>22</td>
<td>10</td>
<td>78</td>
<td>61,682</td>
</tr>
</tbody>
</table>

* When checked against preservation activity of The British Library, CLOCKSS, e-Depot (Netherlands), Global LOCKSS Network, HathiTrust, National Science Library of China and Portico, as reported into The Keepers Registry.
These statistics ought not to be over-interpreted; the results, in terms of percentages, are based only upon those online serials for which an ISSN was known. Nevertheless, a broad picture emerges: only about one quarter of the titles (having ISSN) of each library’s ‘collection’ are being preserved by one or more of the archival organizations reporting into The Keepers Registry. Conversely, about three-quarters of titles (with ISSN) could be said to be at risk of loss to future generations.

The percentage drops significantly for those e-journals being preserved by three or more archives. Recall also that this analysis is limited to those serials for which the ISSN was known, about half the serial titles listed by each library. There is no knowledge of those without an ISSN.

This gloomy picture gets worse in detailed examination of the volumes of content that were preserved for each given serial title. For this there are no simple statistics to report, as this would need searches to be carried out on each e-journal title and, more fundamentally, there is no ready access to a list of ‘ever issued’ volumes in digital form. The results set for _Serials Review_, shown as Figure 4, gives some illustration. This is found by searching on the title or on the ISSN.

Figure 4: Using The Keepers Registry To Assess Archived Extent (May 2013)

This e-journal is being actively preserved by three different archiving agencies. This itself provides re-assurance. However, what is true for a large publisher is less likely to be true for the long tail of smaller publishers. Moreover, not every issue and volume is held by all. For example, e-Depot seems to be missing v.21 and v.1-6 and v.31 are missing at Portico. The general point is that we need to ensure that all content is being kept for the future.

**Looking Forward with 2020 Vision**

We wish to prompt a long view in order to prompt action in the short and medium term. What would success (or failure) look like in 2020, a date which is either within the working lives of many that are now in senior positions?

By 2020 there is every prospect that the digital format will be even more pervasive than today, with greater dependence upon the Web (or its successor). Will we then have assurance of preservation and of access to scholarly statement and to other resources required for scholarship, scientific progress and well-informed civic society? Can we be confident of foresight and concerted international action in the lead up to a future IFLA Conference in 2020?
Consider alternative news stories in 2020 that might break the month before that meeting of IFLA.

**Good News, July 2020**
World leaders congratulate universities and their libraries because The Answer (to the urgent global challenge of xxx), published online in 2008, is still available for use by today’s policy-makers and entrepreneurs as well as by researchers and students – despite the loss of the original website and publisher and the catastrophic mix of earthquake, terrorist attack, format obsolescence, computer virus and bit rot, that threatened permanent loss of The Answer as it was available only in digital form.

Or

**Bad News, July 2020**
World leaders are dismayed to learn that The Answer (to the urgent global challenge of xxx), published online in 2008, is no longer available for use by today’s policy-makers and entrepreneurs or by researchers and students. Distraught citizens despair of universities for all their fine talk, faculty researchers and students wonder why they had such misplaced confidence in their librarians, and the publisher of that lost content is long gone. Although much was known of format obsolescence, computer virus and bit rot; much was known of terrorist attacks and other criminal activity on the Internet; and the devastating effects of earthquake, flood and fire had been there for all to learn from the days of Alexandria, there had been a failure to act.

So how might we envision a future that we would prefer, and how might we make that a reality? This will only be the case if individuals now in senior position take timely and systematic action – before they hand on the baton or as they rise to lead the profession in 2020. The overall objective is to ensure active and quality-assured archiving of the digital content that librarians think are important for their patrons: for researchers, students, their teachers and for citizens and children, now and into the future.

**Actively Engage Archiving Organisations**

What is plain is that the archiving organisations that have already stepped forward as long-term digital shelving need practical and financial support as well as policy guidance. This needs to come from libraries, whether or not libraries hold digital content themselves, do so as cooperatives or contract with some third party, including a national library. Support also needs to come from publishers of digital content – especially if they do not permit ‘local hosting’ of the e-journal content, including back issues.

This can be summarized in a four-point mantra for libraries and publishers:

1) Assign an identifier at the ‘point of issue’ for a stream of digital content.
   • If it is worth preserving for the long term then it should have an identifier.
2) Ensure that (digital) content is archived routinely.
   • Have others/peers do that for you too; lots of copies keep stuff safe(r).
3) Tell someone what you are doing (and how) and what you hold.
   • So that all can know what has been preserved and what is still at risk of loss.
4) Ensure that the terms of access for the archived content (now and when triggered as orphaned) is fit for purpose and is simply stated.
   • The purpose of preservation today is assurance of access tomorrow.
The only consensus, when planning explicitly for the inevitability of failure, is that “digital information is best preserved by replicating it at multiple archives run by autonomous organizations”. Support should be given to several organizations as a matter of good preservation policy, in much the way that that there was re-assurance that printed copies of literature were held in multiple libraries: it might have been good policy for the Alexandria Library to have made copies and store them securely elsewhere.

**Actively Monitor Progress**

Not only do libraries need to require publishers to engage with archiving agencies but they also need to monitor what is actually being archived, prompting publishers and archiving agencies to be alert to what is missing - at the title-level and in terms of the extent to which all issues and volumes have been successfully ingested.

Clearly we mean to argue support for The Keepers Registry, as a key tool for successful monitoring of what is being done. During 2012 and into 2013 we have focused our attention so that we can be confident that The Keepers Registry is:

a) “service ready”
   - with good processes & documentation
b) “known & wanted”
   - by defined use communities.
c) “assured of sustainability”
   - for a sensible & agreed time period

We are confident of the first objective (a) and have a stable technological platform that is capable of extensibility and will scale as usage grows. The Keepers Registry gains direct leverage from the ISSN Register, and the pioneering use of the ISSN-L, for metadata on what is to be kept; it engages an international range of archiving organizations, including both research and national libraries, for metadata on what is actually being kept. We are also investing in interoperability processes for the automated ingest of metadata from the Keepers as these grow in number. In many respects it is not only the first such global registry but is especially suited for the international task.

The Registry has been operating successfully as a Beta service since 2011 and at the time of writing we are preparing to move into a more regular service mode. This will include a relaunch with an enlarged set of services, including semi-automated tools to ascertain the archival status of a library’s ‘holdings’.

We are also confident that the Registry can meet its second objective (b) of being known and wanted, and we seek further engagement through this document with the various communities that would wish to use The Keepers Registry. We invite comment on the scope of digital material in focus, especially as the range of ‘scholarly statements’ widens beyond ‘e-journals’ given the use of the Web in scholarly communication. There is something to be said for keeping within the mission of the ISSN Network with respect to the information objects.

---

Terms like ‘electronic archives’ and ‘organisations with archival intent over the long term’ have been used to define who should be eligible as Keepers reporting into the Registry. The Registry is a lens onto self-stated archival activity, and not an audit and certification service, although the Registry should have metadata fields that allow that to be reported. There is also prospect of using the Registry to carry out spot checks on the actuality and completeness of archived issues and volumes for any given e-journal and archive. However, in all respects the Registry must be seen by the archiving organizations as being in their own interests and as consistent with their missions – for this reason we are planning services within a Keepers Area, to promote the idea of a ‘safe places network’.

We must also be assured of sustainability, objective (c), especially given that our task is to monitor what is being done for the long term, hence focus on the period up to and beyond 2020. There is assurance of organizational stability derived from long-lived organizations with established experience and expertise in serials in the field of scholarly communication and in online service delivery:

(i) the ISSN IC operates under UNESCO mandate and coordinates the ISO-based registration of serials across over 80 countries

(ii) EDINA (est. 1996) is designated by Jisc to delivers online services to universities and colleges in the UK and beyond. It builds on the Data Library (est. 1984) as part of the University of Edinburgh (est. 1583).

Financially, we have gained an understanding of the cost-base for the Registry and how to manage that both in organizationally and technologically. On the matter of a revenue stream to cover costs and to assist development, Jisc has confirmed its commitment to underpin The Keepers Registry, having provided the project funding.

In many ways, sustainability can be conflated with governance, especially for a facility that really must be international, be responsive to the library community and command respect from both publishers and intermediaries, and of course the archiving organisations.

Postscript on improvements needed in metadata

There is more that could be said on what in the Abstract we noted was a related priority to tackle the variety of ‘serial issues’ that can improve the effectiveness of archiving and monitoring. This was the second of the three stories told in The Tales from The Keepers Registry: Serial Issues About Archiving & the Web (Burnhill, 2013). These are matters known to cataloguers of serials but are now seen as metadata for archiving organizations. This also includes a data model for serials that proposes a minimally sufficient set of elements of which good metadata to ensure monitoring of what is being successfully preserved.
These technical matters include identification (e.g. ISSN and ISSN-L) of all types of continuing resources, particularly journals but also ongoing ‘integrating resources’ such as databases and Web sites. This of course is the specialist area for the ISSN Network and the ISSN IC, with important information about the ISSN and electronic publications on its website, http://www.issn.org/2-226838-ISSN-and-electronic-publications.php.

Other topics include the continuing need for a universal holdings statement for assurance that each and every volume and issue has been successfully archived. The ONIX for Preservation Holdings standard (http://www.editeur.org/127/ONIX-PH/) has been devised as a standard for notification and exchange of such metadata. Archiving organizations are encouraged to make files available in this format. Then there is the matter of the inconsistent naming of publishers with a new scheme of identifiers (http://isni.org) and that the history of a given serial often involves a change of publisher – see the TRANSFER Code of Practice for publishers to apply whenever a journal is transferred to another publisher (http://www.uksg.org/transfer/).

Postscript on preserving more than the traditional e-journal

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. The third story in The Tales article (Burnhill, 2013) looks beyond e-journals to new research objects and the dynamics of the Web, to the role of citation and fixity, and to broader matters of digital preservation. The story reflects upon seriality, as the Web becomes the principal arena and medium for scholarly discourse. Scientific discourse is now resident on the Web. Much that is issued on the Web is issued nowhere else. Scholarly statement and government report contain data and multimedia. This is especially important when considering what is the copy of record.

Moreover, 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 on the Web changes. Using Memento 17 which enables ‘travel back in time’ by searching the Internet Archive, a recent

---

17 “Memento,” http://www.mementoweb.org/about/
study\textsuperscript{18} reviewed articles in two scholarly repositories in order to establish which cited resources were still current and which were not, and what that had been archived. The University of Edinburgh, including EDINA, have been funded by the Andrew Mellon Foundation 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, in a project called Hiberlink, see http://edina.ac.uk/projects/time-travel_summary.html.

\textsuperscript{18} Robert Sanderson, Mark Phillips and Herbert Van de Sompel “Analyzing the Persistence of Referenced Web Resources with Memento” Submitted to arXiv on 17 May 2011) arXiv:1105.3459v1