Preserving Open Access Journals: A Literature Review

Sigi Jottkandt

Introduction

The move to put scholarly materials freely online has gained momentum over the past decade as scholars confront the realities of what has been called the “crisis in scholarly communications.” This refers to the double pincer movement of skyrocketing serials prices and the systematic contraction in library budgets over the last 30 years. The result is that faculty at all but the most well-endowed institutions are finding it increasingly difficult to access the materials they need for their research. In response, academics are developing strategies to try to mitigate this crisis, including what has become known as the Open Access movement. Open Access is a grassroots attempt by scholars to make their work freely available through the Internet. Peter Suber, one of the leading thinkers of this movement has defined Open Access as “Putting peer reviewed scholarship online. Making it free of most licensing and copyright restrictions. Removing the barriers to serious research.”

Pioneered by the STM communities, the practice of self-archiving (“Green” Open Access) has been encouraged by libraries and university administrations. Since the early 2000s, a large network of library-based institutional repositories has been implemented worldwide in response to growing numbers of university open access mandates. Open access publishing, the publication of free, “born digital” scholarly materials under Open Access licenses (“Gold” Open Access), has also recently begun to make important strides, as university presses and even some commercial publishers such as Springer experiment with alternative business models that enable them to offer free access to readers.

However, the majority of current open access publishing efforts is conducted by not by publishers but by scholars themselves in the form of self-published online journals and scholarly websites. The Directory of Open Access Journals lists around 4000 peer-reviewed, scholarly journals whose contents are immediately and freely accessible online.3

**Preservation of Open Access Materials**

Open access materials such as online journals and scholarly websites are particularly at risk of disappearing, although they possess many features that should in principle make them very attractive for preservation efforts. First, as Clifford Lynch has pointed out, clearing property rights is one of the most difficult problems confronting preservation efforts.4 Covey et al. have shown that obtaining publisher permission to preserve materials can be immensely time-consuming, and even gaining simple clarity of ownership and rights can be extremely difficult as publishers consolidate, go out of business or fail to keep adequate records.5 Yet obtaining permission to preserve materials is critical to any preservation effort. In the words of the Blue Ribbon Taskforce’s Interim Report, “Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation,” “no responsible institution will preserve materials to which it has no legal right.”6 To the extent that they are published under Open Access licenses that stipulate the kinds of uses that are permitted, OA materials require no additional permission in order to be archived and preserved. They thus eliminate a significant, and frequently difficult, part of the preservation workflow.

---


Another issue related to efforts to preserve toll-access material is also irrelevant to Open Access materials. This is the question of access to the preserved materials. Covey et al indicate that a significant workload is attached to negotiating access agreements with publishers. Frequently, the conditions under which preservation is granted is the restriction or, in the case of “dark” archives, the complete prohibition of access, barring certain “trigger” events which permit content to be released. Identifying and agreeing on such “triggers” consumes significant resources. By contrast, Open Access materials are by definition open to everyone, meaning that no extra resources need to be devoted to preventing access and/or maintaining “moving walls” of content at specified dates. Open Access thus reduces one of the key costs of preservation, namely, employee time. As an additional benefit, the OA community’s use of the Creative Commons licensing framework sheds some much-needed light in the area of electronic rights, which as the Blue Ribbon Taskforce Report comments, are “sometimes obscure in the digital realm.”

Finally, the use of open standards in Open Access journal publishing software and the OA community’s promotion of open formats such as odt and XML assist with preservation efforts. Open standards are technical specifications that have been developed according to a transparent and participatory process and made publicly available. Open standards assist with both of the main approaches to preservation - forward migration and emulation - by providing developers of preservation tools with the blueprints of the original software specifications. Several publishing software systems employing open standards have been developed by the OA publishing community, including TOPAZ, developed by the Public Library of Science for publishing their 7 journals. The Public Knowledge Project’s Open Journal Systems software (OJS) is another publishing system based on open standards and released as open source software. The Public Knowledge Project

reports that to their knowledge there are over 5700 OJS installs worldwide.⁹ Developed by a library consortium based at Simon Fraser University in Canada, OJS conforms to the OAI-PMH protocol, enabling journal contents to be harvested by the OAI and pushed out to search engines and WorldCat. Furthermore, with inbuilt options for indexing with the Directory of Open Access Journals and listing with the preservation effort LOCKSS, OJS automatically handles certain key parts of the workflow that would otherwise require (costly) manual intervention.

In sum, the technical and rights management advantages associated with OA materials should make this form of content attractive to preservation initiatives. Yet, for a number of reasons, OA scholarly materials, particularly the born-digital publications known as “Gold” OA, appear at risk of falling outside existing efforts. A closer examination of a range of approaches to e-journal preservation will help to identify the barriers currently preventing OA materials from being included in these efforts. The subset of OA materials under consideration here are small, independently published, scholar-edited electronic journals. The following literature review covers the existing efforts by the three main actors involved the preservation of electronic journals to determine the extent to which OA journals may qualify for inclusion in these projects. The review therefore covers current efforts by national deposit libraries, third party initiatives, and collaborative preservation initiatives. Given that the OA materials under consideration here are explicitly non-commercial titles, for-profit publishers’ e-journal preservation initiatives are outside the scope of this review.

---

National deposit libraries

National deposit libraries have traditionally been active in the preservation of print materials published within the borders of the country and deposited, in fulfillment of that country’s legal deposit laws, in the national library archives. With the rise of electronic publications, this responsibility for preserving the nation’s heritage has begun to expand to include digital materials. Britain’s new legal deposit legislation passed in 2003, for example, prompted the British Library to form a subcommittee to address issues arising from the deposit of e-journals. In 2007, the British Library’s Digital Preservation Team released a document detailing their preservation plan for e-journal content. This report was the result of a pilot project set up to preserve e-journals in 2006 which included 20 commercial, university, society, and small presses, numbering approximately 200 titles. The BL’s Preservation Plan states that at the time of publication, no final decisions had yet been made regarding the extent of the content to be ingested from publishers. It is therefore not clear first, whether the Preservation Plan relates only to content considered to be of national significance exclusively to the UK, and second, whether the initiative is restricted to commercially published e-journals or if independent Open Access titles can be included in the initiative.

Greater clarity about the selection of online materials for preservation is available from one of the earliest web preservation programs, PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) based at the National Library of Australia. Established in 1996, PANDORA is a consortium of Australian State Libraries who are responsible for selecting online publications and websites that they consider to have national significance and possess long-term


12 Wheatley and McLeod, 3.
research value.\textsuperscript{13} E-Journals are named as one of the selection priorities in the guidelines document, Collecting Australian Online Publications.\textsuperscript{14} Although not mentioned specifically as a collecting priority, a number of Australian Open Access journals (journals edited by Australian academics, and associated with Australian higher education institutions) have been represented in the PANDORA archive since the late 1990s, including the journals, \textit{Cosmos and History} and \textit{Fibreculture}.

A different approach is being taken by the national libraries in Sweden, Finland, Norway and Iceland. These countries have instituted “whole of domain” preservation initiatives. This approach automatically includes everything from the above-named countries whose URIs specify a country domain. For sites that use generic domain names such as .org or .net, there are still opportunities for inclusion if the content is identified as originating in, or concerning subject matter of relevance to the country.

A combination of selective and whole of domain approaches is taken by the Bibliothèque National de France in collaboration with close to 40 national libraries and archives around the world through the International Internet Preservation Consortium.\textsuperscript{15} A registry of 20 of these national archives provides a single point of access to the archived resources of the participating countries. Harvesting methods differ between the countries, and may be either selective, domain, thematic or some variety of all three.

In the absence of electronic deposit legislation, national libraries that pursue whole of domain harvesting to preserve national content may currently offer the best opportunity for the preservation of OA journals whose URIs include country domains or whose content has been identified as originating in or concerning subject matter of relevance to a country involved in preservation of


online content. However, for journals originating from countries with no web preservation program, this solution offers little help. Furthermore, the variation among library harvesting methods means that many OA journals are likely fall outside national library preservation initiatives if, for example, the domain name does not reflect a country specification (for whole of domain harvesting), or in the case of thematic or selective harvesting methods, their content falls outside library selection criteria.

**Third-party initiatives**

Third-party preservation initiatives are fee-for-service preservation models. These fall into the following main access categories: dark archives, “dim” archives, and moving wall access. One of the best known dark archives is Portico, a not-for-profit preservation service that is available to all libraries and publishers on a tiered-fee basis. An initiative of the Ithaka foundation and JSTOR, Portico has quickly developed a reputation within the archiving and preservation community as a trusted repository.\(^\text{16}\) At the time of its launch in 2006, 100 libraries were participating, a number which has in the meantime grown to 657. Portico’s reputation within the publishing community is also fairly high. A 2008 survey of ALPSP publishers showed that 53% of respondents were familiar with Portico’s services.\(^\text{17}\) According to the website, 94 publishers are purchasing Portico’s preservation services, which includes a mechanism for providing subscriber libraries with post-cancellation publisher content (“perpetual access”). Prices for these services vary depending on the publisher’s total journal revenues. For publishers whose revenue falls below $250,000, the archiving fees are fairly modest: $250 per year.

---


\(^{\text{17}}\) Sarah Durrant, *Long-term Preservation: Results from a survey investigating preservation strategies amongst ALPSP publisher members* (ALPSP, 2008), 5.
Another third-party dark archive that may present opportunities for preserving Open Access materials is the Digital Preservation Demonstration Project, Chronopolis.  
Chronopolis is a distributed archive that uses aspects of the existing infrastructure in the educational and research community to provide preservation services for certain collections. A multi-membership partnership funded by the Library of Congress’ National Digital Information Infrastructure and Preservation Program, Chronopolis was developed by UC San Diego’s Supercomputer Center and Libraries, the National Center for Atmospheric Research and the University of Maryland's Institute for Advanced Computer Studies.

Chronopolis currently preserves special collections from only four institutions (the University of Michigan, the California Digital Library, North Carolina State University and UC San Diego), although the website states that the project plans to branch into a broader-reaching, fee-for-service model. It is not clear whether the Chronopolis service will be open to both library and publisher participants in the same manner as Portico, or just to library members only. If the latter, it is conceivable that Chronopolis could preserve Open Access journals, especially if these are regarded as an integral part of the library’s collection, such as journals that are hosted by the library or edited by a faculty member or the recipient of departmental funding.

Both Portico and Chronopolis are examples of “trustworthy repositories”, that is, sets of archiving services and preservation tools that meet a core group of criteria which was established by working groups at the Center for Research Libraries and OCLC at the turn of this century. These criteria address the stability of an organization’s infrastructure, digital object management and technologies, technical infrastructure and security. Since developing this set of guidelines, OCLC has

---


in the meantime put these into practice and established a digital archive for use by the OCLC library community. Headed by the OCLC Digital Collection and Preservation Services, the Digital Archive provides “bit preservation” of web pages harvested by member libraries. Members also contribute bibliographic records for the digital objects and digital archival links that resolve to the permanent copy of the Web object stored in the OCLC Digital Archive. In addition to web pages, the Digital Archive currently preserves around 4,500 scholarly journals in various subject areas as part of the OCLC Electronic Collections Online initiative. Although not originally intended primarily as a preservation initiative, ECO has successfully negotiated with participating publishers to forward migrate the journals in its collection.20

As these examples of third party initiatives demonstrate, OA journals are heavily dependent on library selection decisions when it comes to preservation. With the exception of Portico, which is open to library and publisher members, the third-party preservation initiatives described here are primarily intended to preserve library holdings. Accordingly, to be eligible for third-party preservation, OA journals must either pay for the preservation themselves, or aim to become an integral part of a participating library’s holdings and thereby part of a preservation plan. Portico’s $250USD a year is not a significant amount of money but even this level of outgoing may well be beyond the reach of many editors of OA journals who typically donate their labor but may otherwise have little to no access to funding.21 The other option, to form a relationship with a library through a hosting arrangement for example, can help to defray costs but may at the same time introduce other constraints that may not necessarily contribute to the journal’s scholarly aims, such as a requirement that editors be members of the university faculty.


Collaborative models

Collaborative preservation models are primarily distinguished from the third-party initiatives described above by their absence of fees for participants. Two of the best known collaborative preservation efforts are LOCKSS and CLOCKSS. These are based on open source caching software systems that enable libraries to store copies of journal content on local servers known as LOCKSS boxes. LOCKSS is an acronym for Lots of Copies Keeps Stuff Safe. Like Chronopolis, LOCKSS is a distributed preservation solution, based on the idea that if a single library’s collection is damaged or destroyed, mirrored sites around the world offer a safety net and opportunity for recovering the lost materials. Policies for access to the preserved materials vary. CLOCKSS (Controlled LOCKSS) provides dark archiving of preserved content, while LOCKSS provides access according to participating libraries’ subscriptions.

Decisions concerning which content to preserve is made by the participating libraries who select from a list of publisher titles on the LOCKSS and CLOCKSS web sites. Publishers who wish to have their content preserved through LOCKSS must first establish a LOCKSS consortium for their titles by contacting a participating library and requesting preservation. Once an agreement has been reached, the publisher posts a permission statement (“publisher manifest page”) on the journal’s website and informs the LOCKSS team of the URL. LOCKSS then lists that content as available for preservation. On a recent count, there were at least two open access journals listed as available for preservation in LOCKSS (Other Voices and Vectors), suggesting that LOCKSS offers a good opportunity for small publishers, independent and open access journals to benefit from library preservation projects. Speaking from personal experience, however, in practice the process does not always succeed in listing eligible publishers or journals.
Subject-specific and institutional repositories preservation initiatives

LOCKSS and CLOCKSS are intended to preserve materials from any discipline. In addition to these wide-ranging initiatives, preservation projects have been set up to preserve materials in specific subject areas. PubMed Central, for example, is an extensive archive of open access materials in medical fields. The Institut de l’Information Scientifique of the National Center for Scientific Research (CNRS) collects and preserves French scientific output funded by CNRS. ArXiv, the open access repository for scholarly articles in physics and related fields has recently been “adopted” by 7 university libraries to assist with the maintenance of the archive and long-term preservation.

With their focus on non-commercial digital objects, subject and institutional repositories such as these ought to present excellent opportunities for preserving OA content. However, a 2004 survey by Muir and Ayre found that few of the libraries with digital collections were undertaking preservation activities or had preservation policies.22

Muir and Ayre’s findings are confirmed by a recent Ithaka report which found that a significant majority (66%) of US university libraries are not yet participating in an e-journal preservation initiative. Yet this does not mean that library directors are unaware of the issues, as research into library preservation activities in North America, Australian and Europe conducted by Cloonan and Sanett has shown. Their 2005 study indicated that an increasing number of institutions “are developing policies or seeing the need to develop them” while cost issues are also starting being addressed.23 A case in point is the University of Pennsylvania. John Ockerbloom has described their efforts to create a shared electronic archive along the model of JSTOR/Portico and issued a call for

---


library partners in the project as early as 2002. Nevertheless, without greater consensus on the question of who takes responsibility for preserving electronic scholarly content, there is an ongoing concern that initiatives will continue to emerge on an ad hoc basis of informal distributed networks. For this reason, Ockerbloom recommends establishing a registry of preservation initiatives to keep track of which journals are being preserved and the actors involved.

**Combined models**

This review of existing preservation efforts carried out by the three main participants in preservation activities, national libraries, third party initiatives and collaborative projects, indicates that the majority of existing preservation efforts are targeted at library holdings. This reflects the unfortunate reality, supported by a recent survey of ALPSP publishers, that content producers and rights holders have not yet begun to adequately address the preservation of their electronic content. Through a concern about the loss of cultural heritage and for the sake of the public good, national libraries and archiving bodies have stepped into this gap, creating guidelines for preservation, and implementing web preservation strategies. Unfortunately for smaller publishers and for publishers of Open Access materials, the majority of these initiatives are aimed, either explicitly or implicitly, at preserving commercially published titles. Indeed, a frequently voiced concern of libraries involved in a JISC-funded pilot project for a UK LOCKSS consortium was the lack of “killer content” and the reluctance of NESLi2 publishers to make their titles available for preservation. The same report

---


25 Ockerbloom.

noted, however, that collaborative solutions such as LOCKSS were particularly appropriate for smaller publishers who would otherwise have few options, and for open content. One result is the Open LOCKSS project based at the University of Glasgow, which collects and preserves exclusively Open Access journals.

In addition to these measures, several national libraries have instituted agreements with selected commercial publishers to collect and preserve electronic publications. A pioneer in this area is the Koninklijke Bibliotheek (KB) of the The Netherlands. In 2002, it began operating a storage and preservation facility called the e-Depot. According to a report published in 2004, the e-Depot has agreements to undertake the long-term preservation of Elsevier, Kluwer and the Open Access publisher, BioMed Central’s electronic publications. In 2010, the KB announced its intention to begin archiving and preserving a selection of Open Access journals listed in the Directory of Open Access Journals. Beginning with a pilot project, the KB aims to eventually archive and preserve all the journals in the directory whose content is published in PDF format.

Together, Open LOCKSS, the KB and national web archiving initiatives represent the best opportunities for OA journals from all countries and in all disciplines to be preserved. These initiatives do, however, require active involvement by journal publishers and editors to ensure that their materials are collected. In the first example, this is achieved by posting publisher manifest pages and informing the Open LOCKSS group about the content. In the case of the KB archiving, OA journal editors must make sure their journals are listed with the DOAJ, provide metadata and publish in PDF-A format. In the case of national web archiving initiatives, OA editors and publishers need to ensure that their national library is aware of their content. For selective and thematic preservation programs, editors should contact their national library and request consideration. For countries embarking on whole of domain harvesting, journals should consider using domain names that include country codes (or make their content known to their national library).
Conclusion

Beyond the efforts detailed above, probably the best hope for the future preservation of OA scholarly materials lies in the attitudes of the wider library community itself. OA journals need first to be recognized in the same way as commercial publications for their intrinsic value to research (rather than for extrinsic reasons such as national origin or local faculty involvement). They then should be considered for selection and integrated into a library’s electronic holdings, either through the OPAC, A-Z lists or other library pathways. Ultimately, what is needed is a reassessment of the distinction between paid and free content as the definition of a library’s “holdings”. At that point, the prospects for the long-term preservation of OA content should improve.
Bibliography


http://find.galegroup.com.helicon.vuw.ac.nz/gtx/retrieve.do?contentSet=IAC-Documents&resultListType=RESULT_LIST&qrySerId=Locale%28en%2C%2C%29%3AFQE%3D%28au%2CNone%2C15%29dollar%2C+charles%24&sgHitCountType=None&inPS=true&sort=DateDescend&searchType=AdvancedSearchForm&tabId=T002&prodId=AONE&searchId=R1&currentPosition=1&userGroupName=vuw&docId=A59488495&docType=IAC.


http://eprints.relis.org/archive/00007354/.


Thibodeau, Kenneth. “The Electronic Records Archives Program at the National Archives and Records Administration.” *First Monday* 12, no. 7 (July 2, 2007).


