

**Columbia University Libraries/Information Services  
Strategies for Expanding e-Journal Preservation  
Final Report  
to the Andrew W. Mellon Foundation**

Principal Investigator: Robert Wolven  
Associate University Librarian for  
Bibliographic Services and Collection Development  
Columbia University

Grant Reference Number: 31300635

Grant Period: November 1, 2013 – July 31, 2015

Reporting Period: November 1, 2013 – July 31, 2015

## **Summary**

Columbia University Libraries/Information Services, in collaboration with the Cornell University Library (jointly referred to as 2CUL) has completed an 18 month project intended to expand significantly the preservation coverage of e-journals and to implement strategies that will sustain the initiative beyond the Project. Some expectations and outcomes have been modified during the course of the project in order to meet the underlying objectives of the proposal. During the course of the project, the project team addressed a number of challenges, and identified areas for future development.

## **Objectives: Description of the Project and Purpose of the Grant**

In October 2013, the Andrew W. Mellon foundation awarded a grant of \$150,000 to 2CUL for use over 18 months (later extended to 21 months) to support the expansion of preservation coverage of e-journals and the identification of appropriate paths to enable libraries to secure the preservation of additional material considered priority content. Methods were developed and tested to facilitate the continued expansion of e-journal preservation through ongoing assessment of priorities and documented practices to encourage and expand community engagement. Preservation strategies best suited to diverse types of e-journal publication were identified, pursued, and promoted. Quantitative analyses were undertaken at the beginning and end of the project period to assess the impact of the project and changes in the state of preservation for a defined set of e-journals.

## **Deliverables: Expected Outcomes and Benefits**

The intended outcomes of the project were three-fold: first, to increase the number of preserved e-journals by 1,500-2,000 through direct action and initiate methods and procedures that would preserve a total of 5,000-7,000 titles within two years; second, to apply qualitative and quantitative analyses to characterize the range of non-preserved titles and, in consultation with other libraries and preservation agencies, to identify high-priority content within this group; third, to develop specific, concrete procedures that would serve to sustain and expand the project's accomplishments following its completion.

## **Accomplishments: Summary of Activities and Progress Achieved**

At the outset of the project two groups were established to advise on priorities and methods and to monitor and evaluate progress: a 2CUL group of librarians representing e-resource management, collection development, preservation, and library technology; and the Collection Development Associate University Librarians in the Ivy Plus consortium (a group of twelve libraries including the eight Ivy League institutions plus Duke, Johns Hopkins, MIT, and the University of Chicago). Recognizing that preservation status is not a binary value but covers a range of conditions, we worked with these groups to define a set of preservation statuses: preserved, protected, in process, and not preserved (see Appendix A.) We adapted our methods in working with different types of content, with the objective of achieving the highest status feasible over the course of the project. We also worked with these groups to validate priorities, including the decision to exclude several categories of e-journals that were either considered low priority for preservation, or as offering little chance of success within the project's duration (see Appendix B). Realizing that the corpus of un-preserved e-journals is large, we wanted to characterize what is considered most valuable from a scholarly perspective and thus poses a greater risk of loss.

Throughout the project we also met and consulted with the major parties currently engaged in e-journal preservation, including CLOCKSS, LOCKSS, Portico, the Keepers Registry, the Directory of Open Access Journals (DOAJ), the Public Knowledge Project (PKP), and the Center for Research Libraries. In addition, careful attention was given to outreach efforts to expand the understanding of issues around preservation for e-journals, as a first step towards expanding the number and range of active participants in this work.

The project focused on five main categories in an effort to develop methods for expanding the coverage of e-journal preservation: content produced or made available by major publishers, small publishers, aggregators, universities acting as publishers, and open access journals. Taken together, these categories gave us a target set of roughly 30,000 active e-journal titles in need of preservation.

### *Major publishers*

Title lists from nine major publishers were evaluated for the preservation coverage of their holdings. The Keepers Registry and data provided by Portico were integral in determining the

preservation status of these titles. Shannon Regan, the Project Librarian, contacted seven publishers with at least one non-preserved title to request further information and to encourage steps to complete preservation. Five of the seven provided detailed responses. In general, these publishers deliberately excluded certain categories from third-party preservation initiatives, mainly: indexing and abstracting titles, newsletters, and continuing education e-journals. Two publishers did not respond to repeated inquiries; however, subsequent analysis revealed that virtually all of their titles in question were later committed to preservation through Portico, though we are unable to say whether our inquiries prompted that action.

Through this analysis two additional factors relevant to the preservation of e-journals from major publishers became evident. First, the extent of preservation coverage for major vendor backfiles varies considerably. Second, the transfer of titles from publisher to publisher can have an adverse effect on the preservation of a title, even if both publishers involved participate in a third party preservation initiative.

#### *Small and mid-size Publishers*

In consultation with the Ivy Plus advisory group and with Portico staff, we identified 50 small- to mid-size publishers for detailed analysis and action: 30 already participating in Portico, CLOCKSS, or LOCKSS to some extent, and 20 not participating. Ms Regan again analyzed the preservation status of their e-journals and contacted publishers with non-preserved titles. The obstacles to preservation for small and society presses are in stark contrast to those of major vendors. Small and society publishers face significant cost barriers to participation in third party preservation initiatives. Similarly, most do not have the technological expertise or funds to implement new technologies supporting preservation strategies. Most evident in conversation with small and society presses is a clear lack of understanding as to what digital preservation is and how it may be accomplished. Overwhelmingly, the majority of small and society publishers contacted in conjunction with this project did not participate in any preservation initiative simply because they were not aware of the initiative or did not understand how it worked.

Of the 20 non-participating publishers in this group, two have now joined Portico (with 51 titles preserved or queued for preservation) and two others have been purchased by participating major publishers.

#### *Aggregators*

During the second quarter of the project we initiated discussions with two major aggregators of e-journals, EBSCO Information Services and ProQuest, regarding their potential role in preservation and the barriers to preserving the e-journal content they provide. These discussions continued throughout the course of the project. With EBSCO we entered into a deeper collaboration to explore potential business models for supporting a preservation service and to develop workflows for obtaining the rights to preserve the content distributed by the aggregator.

Through this collaboration Ms Regan contacted over 350 publishers. Most of these publishers publish one or two titles. Overall, there was a 30% response rate in which over half of the

respondents agreed to participate in the pilot preservation service. One of the key takeaways from this project is that these small publishers are not resistant to preservation but simply do not know about digital preservation initiatives or the expectations of the library community in regard to the archival status of journals subscribed. Outreach and education have the potential to go a long way in securing preservation, but this is very labor-intensive work, requiring repeated conversation and explanation to preserve a relatively small number of titles.

In order to assess interest in an aggregator-provided preservation service, EBSCO and staff from our project held discussions with EBSCO's Academic Library Advisory Group and with the Ivy Plus Collection Development Group. As of this writing EBSCO has chosen not to pursue offering preservation as a service, but has begun to include the right to preserve content in future agreements with publishers, and to pass that content onto a third-party preservation agency under defined conditions. They are continuing to explore ways in which to act on these agreements so that the content which they distribute is preserved.

Our discussion with Proquest resulted with a different outcome. The Proquest staff involved in the conversations with the 2CUL team were intrigued with the concept of "preservation." They decided to develop and administer a survey to assess the opinions and expectations of member libraries and publishers. They expect to complete and administer this survey later this year. Another related discussion was the pilot work CRL and Proquest have been involved in (within the context of the CRL e-newspaper preservation project) to experiment with a light-TRAC process. The purpose of this concept is to assess publishers/aggregators' technical infrastructures, policies, work flows, and business models in order to assess their preparedness and reliability for performing preservation responsibilities.

### *Open Access*

Freely accessible e-journals comprise the largest, most diverse, and in all likelihood most problematic category for preservation in our target set. They vary widely in importance, in content, and in publication methods and source. There is considerable overlap with other categories addressed in the project, and we excluded from this group open access titles provided through major publishers and aggregators. There are also other active preservation efforts focused on specific groups of open access e-journals. LOCKSS has developed a plug-in that works with the latest version of the Open Journal System (OJS) platform and the Public Knowledge Project is in the process of formalizing their private LOCKSS network. In addition, Portico has a helpful export plugin for OJS users to ensure Portico receives the most accurate files. After consultation with the directors of LOCKSS, the DOAJ, and the PKP program, we decided to avoid overlap with their work and focus on small, independently published titles.

After further conversation with the directors of LOCKSS and CLOCKSS, we decided that the most effective method for immediate action would be to capture the e-journal content through web archiving, with the aim of using either a Private LOCKSS Network or a Fedora repository for long-term preservation. Through its pre-existing web archiving program, Columbia has already preserved 77 e-serials in the Human Rights and Avery Library collections. To test this approach further we selected 77 titles that had been identified as important by selectors within

2CUL. Ms Regan contacted the publishers with details regarding the project and proposed the use of web archiving tools for e-journal preservation. The response was swift and overwhelmingly positive, only one publisher asked to be excluded. Columbia's web archiving team then harvested the content and analyzed the results to identify and resolve issues of scope and quality control. (The collection is available at <https://www.archive-it.org/collections/5921>) We believe this method can easily be extended to a large number of freely accessible e-journals. Cornell is in the process of considering a collaborative project with the Committee on Research Materials on Southeast Asia (CORMOSEA) to focus on the future of open access journals from this SEA region. This nascent vision is considered a component of the SEA newspaper project led by Cornell.

### *University Publications*

A relatively small but growing number of e-journals are published by universities, often through units associated with their libraries. Many are published through the OJS platform noted above, though not all are open access. We decided to devote special attention to this category, recognizing that publishers in this group often have active digital preservation programs whose content is not tracked by Keepers Registry.

We met with the director of Columbia's Center for Digital Research and Scholarship (CDRS) to better understand issues affecting preservation of the 20 journals published by this unit (<http://cdrs.columbia.edu/cdrsmain/projects/#journals>). We examined the websites of a number of other university publishers for statements about archiving practice. Ms Regan contacted a few of the major publishers directly, to inquire about their preservation practices and policies. We found that most of these publishers have not made formal, open preservation commitments, but do plan to preserve their e-journal content.

### **Impact of the Project**

One major objective of the project was to secure the preservation of 5,000 e-journal titles that had not been preserved at the project's outset. It is impossible to say with any precision how many e-journals have been preserved as a direct result of this project. As an example of the complexity involved, one major publisher did not respond to our initiative identifying titles among their online publications that were not preserved; shortly after, however, this publisher deposited over 200 of these titles with Portico. Similarly, Portico engaged directly with some smaller and mid-sized publishers based on our identification of high-priority titles. At least one university library decided to assume archiving responsibility for the journals it publishes after discussion with our project staff. Other titles have made progress towards preservation as a result of our work, but are not yet fully preserved. These include titles from publishers who have now granted archiving rights to an aggregator, and titles in Columbia's web archive; we would consider these to be in "protected" status.

It is even more difficult to measure progress toward a goal of "full" preservation. As we have repeatedly noted, the definition of e-journals (not to mention preservation) is imprecise. We selected e-journals in Columbia's and Cornell's catalogs as a corpus to analyze and focus on. In

the course of our work, however, we encountered and preserved titles that had not been cataloged, while the total number of e-journals added to our catalogs over the past two years appears to exceed the number preserved during the course of the project.

While we can't directly link cause and effect, we were able to analyze the net change in preservation status since the project's beginning for 58,000+ e-journals held by Columbia as of July 2013. This data set was selected for analysis because it had been analyzed by Keepers Registry just prior to the start of the project. In August 2013, 44,889 titles out of 58,556 had not been preserved by any agency tracked by Keepers Registry. At the end of June 2015, 4,043 titles in this set had been preserved by at least one agency. Thus, the percentage of preserved titles increased from 23.3% to 30.2%. The situation improves further when certain exclusion categories are removed from the list. For example, if Chinese, Japanese, and Korean titles are excluded, 34% of the remaining titles are preserved. Similar analysis of three sets of data from Cornell showed an increase of at least 4,570 in the number of preserved titles, with 31.8% of the full set of titles currently preserved.

Beyond the information tracked by Keepers Registry, the impact of the project has been smaller in numbers, but perhaps greater in effect. One major aggregator is now acquiring preservation rights in all of its licenses with publishers. Web archiving has been tested and proven feasible as a means of preserving journals that are unlikely candidates for CLOCKSS and Portico. Recognition of the complexity and diversity of the preservation landscape will increase opportunities for action by more parties, resulting in protection of more of the total e-journal corpus.

## **Challenges**

Several challenges were encountered during the course of the project, and though most had been anticipated, they did require frequent adjustments in the detailed work plan. Perhaps the biggest challenge was simply the time required to explain the purpose of the project, including libraries' expectations and needs regarding preservation of e-journals, to many parties with diverse backgrounds and perspectives. Publishers, editors, and aggregators each had different degrees of awareness of issues, but also different understanding of the meaning of terms such as "preservation" and "archiving." All of those working on the project spent considerable time answering questions. This became somewhat easier as we refined scripts for communicating, developed FAQs, and clarified the language in proposed licenses, but the process of negotiating preservation agreements remains labor intensive.

Adding to this challenge was the fact that preservation is not the highest priority for most of the parties we worked with. Staff at Portico and EBSCO were extremely helpful in providing contact information that helped us reach the right people at the publishers involved; even so, capturing and sustaining the attention of those individuals was difficult. The same proved true in attempting to work with staff at the two aggregators; while our interactions were cordial and informative, it was not always easy to get the attention of those needed to resolve issues.

Changes in the e-journal publishing landscape also required adjustments in project plans. Publishers we planned to work with merged with, or were purchased by others. E-journal titles under investigation transferred to new publishers. New preservation initiatives such as the PKP work on OJS titles were begun. DOAJ expanded the journal description data gathered from the OA publishers, now requesting information about the preservation status of journals listed in the directory. In general, these were positive developments for e-journal preservation, but they did cause a few false starts and course corrections in our work.

Perhaps the most surprising and disturbing challenge was the degree of questioning we encountered within the library community itself regarding the importance of taking action to preserve e-journals. This was expressed as a combination of (in our view, misplaced) confidence that publishers and aggregators can be relied on to archive their own content, plus doubts about the technical and economic reliability of existing third-party preservation agencies. While these views were by no means universal, we did find a number of librarians questioning the costs of preservation, when no major losses have occurred to arouse immediate concern. The reluctance from librarians to aggressively pursue e-journal preservation may be influenced by confusion as to where the responsibility for preservation lies; with publishers, third party agencies, or libraries. Actions to further expand e-journal preservation will bring additional costs. Assessment of the economic impact of this work lies outside the scope of the project, but will be an important area for further attention.

At the outset of the project, one of our goals was working with the BorrowDirect Collection Development group to expand the current model license terms to incorporate preservation-related expectations and requirements. It was during this project that with **funding from the Andrew W. Mellon Foundation** the Center for Research Libraries announced the newly revised model license, incorporating the recent best practices. Although preliminary in nature, the new model license incorporates some licensee rights related to third party archiving of licensed journals.

### **Project Personnel: Significant Board, Management, and Staff Changes**

There were no significant changes in the staffing plan during the course of the project. The timeline for the project was extended by three months, due to delay in filling the project librarian position.

The co-Principal Investigators are Bob Wolven and Oya Y. Rieger from Columbia and Cornell universities, respectively. Joyce McDonough is Columbia's Director of Continuing and Electronic Resources Management. The Project Librarian is Shannon Regan.

### **Recent Publications, News Articles, or Other Related Materials**

Details regarding progress on the project were shared via presentations at several library conferences in 2014-2015.

Oya Y. Rieger organized and spoke on a panel at the Fall 2014 Coalition for Networked Information membership meeting regarding the topic, “e-Journal Archiving: Changing Landscape.”

Shannon Regan presented, “Case Study for Expanding e-Journal Preservation: An in Depth Analysis of De Gruyter’s Acquisition of 67 Berkeley Electronic Press Journals,” at the 10<sup>th</sup> Annual Electronic Resources & Libraries Conference in Austin, TX on February 24, 2015. <http://academiccommons.columbia.edu/catalog/ac%3A187118>

Shannon also presented, “Strategies for Expanding e-Journal Preservation,” at the 30<sup>th</sup> Annual NASIG Conference in Arlington, VA on May 28, 2015. In addition, Shannon reprised her NASIG presentation for an audience at the National Library of Medicine on June 15, 2015. <http://academiccommons.columbia.edu/catalog/ac%3A187115>

Oya Y. Rieger & Bob Wolven presented at the Fall 2013 Coalition for Networked Information membership meeting regarding the topic, “Strategies for Expanding e-Journal Preservation” on at Washington, DC.

### **Next Steps and Recommendations for Further Action**

The work completed under this project suggests several promising paths for further action to extend the range and number of preserved e-journals, and to further build community consensus around priorities. The libraries of Cornell and Columbia will pursue some of these actions in the coming months, and will continue working with stakeholders such as LOCKSS, Portico, Keepers Registry, CRL, and the major aggregators to build on this work. The nascent Ivy Plus collection program offers a potential incubator for some initiatives such as web archiving of e-journals, and a platform for publicizing our work. Specific recommendations for future action are included in Appendix C.