

Project Impetus and Strategy

Project Impetus

Evidence shows that the extent of e-journal preservation has not kept pace with the growth of electronic publication. Studies comparing the e-journal holdings of major research libraries with the titles currently preserved by major agencies have consistently found that only 20-25%, at most, of the titles with ISSN's currently collected – let alone published – have been preserved. In early 2011, the libraries of Cornell and Columbia conducted a study as part of the 2CUL collaboration, and found, for example, that LOCKSS and Portico combine to preserve only a relatively small percentage of these libraries' e-journal holdings, less than 15% of Cornell's e-journal titles as a whole.^[1] In the fall of 2012, a study using The Keepers Registry comparing the e-journal holdings of Columbia, Cornell, and Duke with the e-journals preserved by seven different agencies, yielded similar results, showing that only 22-27% of the subset of titles with an assigned ISSN had any volumes archived. Moreover, the extent of volumes archived for any given title varied greatly and was often sparse.^[2]

Beyond the diversity of content, individual libraries, despite their concern for preservation, often lack effective means for taking action. One of the revealing findings of the 2CUL e-journal preservation study was that many staff at Cornell and Columbia only had a superficial understanding of the relevant preservation strategies and their implications – and of the roles of libraries in advancing the e-journal preservation front. Selection and acquisition processes may not involve any direct interaction with the publisher; many titles are acquired as parts of large packages, with no comprehensive provision for preservation. One of the proactive strategies proposed is developing language for a model license addressing preservation and sharing it with the publishing and library communities to set a timeline for implementation (e.g., in five years, all ARL libraries will aim to use the same licensing language). Also identified as critical was mobilizing advocacy to engage the key stakeholders and providing incentives for community-wide benefits. It is important systematically to engage publishers to address the problem and make this process transparent and public.

Strategy

While quantitative goals are difficult to set until methods have been tried and proven, our objective is to secure the preservation of a significant number of e-journals during the course of the project, and to establish methods that can be more broadly adopted following the project's completion, leading to further expansion of preserved e-journals. We will make an effort to preserve 5,000-7,500 additional e-journal titles. Due to the short duration of the proposed project and the need to work with many publishers, each responsible for only a few titles, the number of journals preserved through direct action by the 2CUL team is likely to be lower, in the range of 1,500-2,000 titles. Thus, the objective of the project is to initiate the process and put in place a work plan and timeline to enable the preservation of several thousand high-risk journals. During the first quarter of the project, we will investigate current impediments to securing preservation and explore means of overcoming obstacles. As we go through this applied research process, we will be able to refine our strategy, desired outcomes, and timeline based on the early findings and indicators.

Realizing that the corpus of un-preserved e-journals is large, one of our goals is to characterize what is considered most valuable from a scholarly perspective and thus poses a greater risk of loss. We will use quantitative and qualitative analyses to determine content that requires priority of attention, content that is both important and vulnerable from the perspective of research libraries, in order to guide and assist the effort of archiving agencies vis-à-vis the publishers, or other custodians, of that content.

The 2CUL team will select a set of representative titles from high-priority categories to develop, test, and promote methods to expand the body of preserved e-journal content. In parallel with this work, which aims to increase the number of preserved e-journal titles, we will use the full list of un-preserved titles to assess the volume and vulnerability of content across a range of publication categories (such as peer-reviewed, scholarly journals, trade newsletters, government publications, and annual reports and bulletins of organizations), and assess priorities within those categories. Finally, we will work to broaden the impact of the project's activities by raising awareness of e-journal preservation risks and identifying effective ways for libraries to engage in addressing those challenges. Each of these project strands will be implemented in collaboration with the recently constituted Collection Development group within BorrowDirect (a consortium of the libraries of eight Ivy League universities plus the Massachusetts Institute of Technology and the University of Chicago) in order to test and validate the generalizability of the findings and strategies.

Neither of the organizations had a formal process in place for the identification of e-journals for preservation consideration and, currently, neither organization records and manages details related to potential LOCKSS or Portico access in the libraries' local systems for the acquisition and management of electronic collections. Anecdotal evidence suggests a similar picture for other libraries: a lack of a streamlined process for attending to the archival status of e-journals. As we work to address the archival challenges of traditional e-journals, publishing continues to change rapidly and it includes enriched content that is layered, interactive, and dynamic. We risk falling behind. The success and sustainability of the proposed project requires concerted and continuing support from the key stakeholders such as libraries, publishing agencies, and preservation service providers. Therefore, the strategy described above will be complemented with action that aims to address e-journal preservation upstream as a shared responsibility.

Expected Outcomes

- Identify important and vulnerable e-journal content from the perspective of the research library through a quantitative and qualitative methodology;
- Select a set of representative titles from high-priority categories to develop, test, and promote appropriate archival strategies based on content type and origin;
- Create and test methodologies that will result in preservation of this content;
- Document and share findings to facilitate the continued expansion of e-journal preservation through ongoing assessment of priorities and documented practices;
- Collaborate with the BorrowDirect Collection Development group to validate the assessment of priorities and the selection of titles for initial action, to advise on proposed methods of work, to refine recommendations for best practice and model license terms, and to assess results (to ensure the generalizability of the findings and strategies);
- Work to broaden the impact of the project's activities by raising awareness of e-journal preservation risks and identifying effective ways for libraries to engage in addressing those challenges (to encourage and expand community engagement);
- Document and promote practices designed to expand the number of libraries actively engaged in pursuing and supporting e-journal preservation;
- Engage libraries, publishers, societies, and other key stakeholders in analyzing current impediments to securing preservation agreements and test methods of working with appropriate parties (publishers, professional societies, e-journal aggregators, and preservation agencies) to overcome these obstacles;
- Ensure the sustainability of the proposed project, develop strategies to address e-journal preservation upstream as a shared responsibility (e.g., creation and implementation of model license agreements);
- Create forums for exchanging information about relevant preservation strategies and their implications and the roles of libraries in advancing the e-journal preservation front in order to encourage streamlined processes for attending to the archival status of e-journals.

Our overall goal is to take actions that will secure a preservation strategy for an additional 5,000-7,500 e-journal titles. We aim to have 1,500-2,000 titles preserved through direct action by the project team, with an additional 3,500-6,000 titles preserved through action by preservation agencies, e-journal aggregators, and other libraries, applying methods developed through the project, either within the time frame of the project or in the following one to two years. During the first quarter of project, the Project Advisory Team will select an initial set of 20-25 e-journals for detailed investigation and action. The outcomes of this process will help the 2CUL team further refine and modify our strategy and projections.

By analyzing in detail the preservation status of the e-journal holdings of two major research libraries, the project will provide reliable data on the extent and nature of current gaps, and will promote a methodology for continuing and extending the analysis after the project's completion. The data provided through these comparisons will help the research library community to determine priorities for action using multiple criteria ('Importance', as measured by titles that are most widely held; 'Risk', as measured by titles uniquely held and/or not falling under an existing preservation mandate such as that of the British Library or other national library; etc.).

Using specific examples of titles identified as having high priority for action, the project will also test and promote methods appropriate to different categories of e-journals, to enable libraries, publishers, aggregators, and preservation agencies to take more effective action, including model language to include in licenses, procedures for working with publishers to facilitate adoption of sound preservation practices and recommended actions for preserving content locally. Project results will be broadly disseminated, with the intention of building wider community participation and sustained momentum. The success and sustainability of the proposed project requires concerted and continuing support from the key stakeholders such as libraries, publishing agencies, and preservation service providers.

1. The 2CUL study is available at: <http://2cul.org/sites/default/files/2CULLOCKSSFinalReport.pdf>

2. Burnhill, Peter. "Tales from The Keepers Registry: Serial Issues About Archiving & the Web," *Serials Review*. Volume 39, Issue 1, March 2013, Pages 3--20.

<http://www.sciencedirect.com/science/article/pii/S0098791313000178>