International Print and Digital Repositories
Initiatives in the United States: CRL, Portico, LOCKSS, Internet Archive

Print Archives in the United States
James Simon, Director of International Resources, Center for Research Libraries

Background

A number of pressures on academic libraries imperil the long-term survivability of printed knowledge and heritage materials. Ever-growing volumes of materials, costs of preserving and delivering paper-based research resources, and researchers' growing demand for source materials in electronic formats all produce strain on our institutions.

Libraries have traditionally engaged in processes of selection of material for long-term retention—and, conversely, de-selection of other material. While some activists have charged that libraries have failed in their roles by de-selecting, or "weeding," their collections, factors such as sheer space and finances dictate that libraries cannot retain and preserve every resource acquired.

The development of Print Archives evolved out of a series of discussions in the late 1990's relating to the role of the artifact in library and archival collections, and the value of those materials for scholarship and teaching. The report produced by the Council of Library and Information Resources (CLIR) entitled "**The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections**"[1] lays out the finding of the Task Force on the Artifact in Library Collections and attempts to articulate a framework for making or evaluating institutional policies for the retention of published materials and archival or unpublished materials in their original form.

The task force was charged to answer specifically these questions:
- What qualities of an original are useful or necessary to retain in their original form? Under what circumstances are original materials required for research?
- When is it sufficient and appropriate to capture intellectual content through reformatting and not necessarily retain the original?
- Which preservation options provide the most appropriate and cost-effective means of preserving the original?
- From both custodial and scholarly perspectives, what are the advantages and disadvantages of these various preservation options?

The report deals in detail with the ways in which artifacts and their physical frailties affect their research value—originality, faithfulness, fixity, and stability—over time and with how libraries can minimize the risk of unacceptable loss of that value. It investigates the specific issues around selected media, such as paper-based printed matter, moving image and recorded sound materials, and objects that exist in digital form, and points to two key strategies—*preventive maintenance* and *use of surrogates*—currently in use in libraries that can be scaled up to deal with a problem of this magnitude. Specific recommendations are provided for different media. The recommendations for print included:

---

[1] The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections. Council on Library and Information Resources. November 2001. http://www.clir.org/pubs/reports/pub103/pub103.pdf (accessed 10/7/2006).

- Establishing regional repositories to house and provide proper treatment of low-use print matter. These repositories would provide access to artifacts aggregated from different institutions under terms to be worked out. Such repositories might begin by taking in journals and monographs that are widely available in digital form.
- Convening a national group to investigate the establishment of archival repositories that would retain a "last, best copy" of American imprints.
- Building inter-institutional networks of information sharing about the status of artifacts and delegation of responsibilities for caring for them. These networks would obviate unwanted duplication and encourage libraries to take responsibility for the preservation of their most important or rarely held materials. These networks would include a registry of digitized collections that has information about where the originals are located, who has responsibility for them, and under what terms they may be used.

The task force recommendations called for planned growth of cooperatively managed repositories for little-used materials. This aspect was taken up as the subject of another CLIR report produced by Bernard Reilly, president of the Center for Research Libraries (CRL), entitled "**Developing Print Repositories: Models for Shared Preservation and Access.**"[2] CLIR commissioned CRL to investigate existing models of repositories consortial use as well as models of archival repositories. The report includes information about both regional repositories and those that collect on behalf of an entire nation. The authors of the publication also document features of repositories that have developed concentrations of rare or little-used materials in specific collecting areas, as well as repositories that collect and preserve "last-copy" imprints (that is, items that are rare and possibly unique and thus in need of retention).

Around the same time (July 2003), the Center for Research Libraries organized a conference to address the urgent issues associated with preserving the nation's print materials. The "**Preserving America's Print Resources" (PAPR)**"[3] initiative drew together hundreds of concerned professionals from a cross-section of the library industry to answer the question: "How can the nation's libraries build upon existing regional and national efforts to optimize management of critical knowledge resources in printed form?" The program brought together a slate of presenters from the U.S. and abroad who are actively engaged in major repository and "collection of record" programs. Presenters detailed successes and innovations of existing inter-institutional efforts, and proposed realistic cooperative measures that might build upon those efforts to ensure the long-term survival of printed heritage materials.

Current Print Archive activities in the U.S.

The reports described above paint a broad picture of the concerns and discussions designed to meet the challenges of preservation and retention of legacy materials for future use. There have been a number of moderate successes. The following presents a few of those efforts.

JSTOR Archives

The majority of early print archive efforts have concentrated on the availability of print versions of digitized journals. In the U.S., a number of programs have centered on titles made accessible through JSTOR, the largest not-for-profit archive of important scholarly journals.

Since its inception, JSTOR has acknowledged the need to preserve original paper copies of the journals in the archive. As digital archival standards and computing platforms evolve, the JSTOR archive will need to evolve, as well. Print repositories would maintain the paper artifacts for future consultation and use by JSTOR. Maintaining paper repositories would also provide a fail-safe in case of a disaster in which the digital archive was damaged or degraded. Finally, JSTOR, through cooperative agreements to store its

---

[2] Reilly, Jr., Bernard F. "Developing Print Repositories: Models for Shared Preservation and Access." Council on Library and Information Resources. June 2003. http://www.clir.org/pubs/reports/pub117/pub117.pdf (accessed 10/7/2006).
[3] "Preserving America's Print Resources." Center for Research Libraries. http://www.crl.edu/content.asp?l1=1&l2=54&l3=105 (Accessed 10/7/2006).

material, hoped to provide a practical example of how paper repositories may be compiled, managed, and maintained.[4]

In 2002, JSTOR staff met with members of the library community from around the United States to discuss dark vs. light archives, distributed vs. centralized archives, JSTOR vs. library ownership of the paper volumes, and the logistics and costs of collecting, validating, and storing paper volumes to be included in a repository. The Center for Research Libraries became the first North American repository of copies of every JSTOR journal in its database.

Center for Research Libraries JSTOR Archive[5]

CRL has built its archive of JSTOR journals from deposits and receipts from a number of participating CRL members, including the University of Arizona, Northwestern University, and other libraries throughout the U.S. The goal of the project is to acquire, through donation, a complete file of the print copies of journals available through JSTOR. The archive serves to provide assurance to the CRL members, and to the library community, that paper copies of all JSTOR titles will be available in the future.

Donations are accepted from any institution, CRL members and non-members alike, who are willing to donate materials to this repository. The schedule for donations is flexible, though CRL endeavors to keep current with JSTOR holdings in the online archive, dictated by the "moving wall" of JSTOR collections (typically 1 to 5 years behind the most recently published journal issue). Materials may be offered as a single title, or parts of single titles, multiple titles from a single subject sequence, all titles from a single subject sequence, or multiple titles, or parts of multiple titles, from any, disciplines.

Materials are collated at CRL, and holding records are loaded into CRL's catalog. The materials are intended to be used by CRL members on request, though records are currently suppressed at this time to discourage unwarranted loan requests.

University of California/California Digital Library JSTOR Archive

JSTOR itself has moved to create additional security for its digital collection by signing agreements with university systems to maintain paper repositories of JSTOR titles at those locations. In 2004, the University of California (UC) system signed with JSTOR to maintain the archives at the Southern Regional Library Facility on UCLA's campus, current with JSTOR's moving wall for each title.

Various institutions in the UC system were asked to be the primary contributor for particular subjects, based on existing subject strengths of the respective library. For instance, UC-Irvine was selected to provide titles in the subject areas of Anthropology, Botany, Ecology, Economics, History, Music, and Political Science. Irvine would also be a secondary contributor to fill in holdings of another primary depositor.[6] The repository will contain the full run of each journal, of which UC will retain ownership.

To address access-related concerns, JSTOR agreed that a "very dim," rather than completely dark, archive would be acceptable. Under these conditions, library patrons may only gain access to a volume from the repository if all other practical means of access are deemed unacceptable. In such cases, a volume will be circulated for in-library use only and will undergo a thorough review upon its return and re-deposit into the repository. Although a patron may occasionally

---

[4] Text selected from JSTORNEWS, No. 8, Issue 3, December 2004. http://www.jstor.org/news/2004.12/repositories.html (Accessed 10/7/2006).
[5] "CRL/JSTOR Print Archive Project." http://www.crl.edu/content.asp?l1=13&l2=19&l3=35&l4=62&l5=12 (Accessed 10/7/2006).
[6] "UC Partnership with JSTOR to Create a Print Archive." http://www.lib.uci.edu/libraries/projects/jstor/index.html (Accessed 10/7/2006).

need to consult a paper volume from the repository, it is anticipated that such cases will be very rare.[7]

This agreement supports the libraries' ongoing collaborative efforts to de-duplicate holdings and develop shared print collections, maximizing both scarce shelf space and limited collections budgets.[8]

Five Colleges of Massachusetts

In 1999, the presidents of the Five Colleges of Massachusetts—Amherst, Hampshire, Mount Holyoke, Smith Colleges, and the University of Massachusetts at Amherst—approved a plan to operate a consortial library depository that would have a distinctive mission and governance. They assigned responsibility to Five Colleges, Inc., an existing consortium, to run the depository.

Among the first decisions was to move to storage the journals that are available online, such as those mounted on JSTOR. By moving sets offsite, the Five Colleges would be able to create a complete run of any given journal. The Five Colleges agreed to share ownership of the jointly stored material (with the exception of the University of Massachusetts, a public research institution that is required by mandate to retain ownership of its materials).

This model has shown some success. A similar effort is underway through the CONSORT Colleges of Ohio (Denison University, Kenyon College, Ohio Wesleyan University, and the College of Wooster) to develop centralized offsite storage for journals at the shared CONStor facility.[9]

Distributed Print Archive Project (CRL)[10]

Occurring at the same time as CRL's involvement in the PAPR conference and beginning the JSTOR archive, CRL received a 2-year assessment grant intended to "model and test a framework for the distributed, long-term retention of artifactual collections, using JSTOR journals as a test bed of materials."

The grant supported exploration of four areas:
- o   The economics of distributed, long-term retention of artifactual collections
- o   The necessary framework and logistical support for cooperative print archives and CRL's role in providing this support
- o   The development of formal agreements on specific terms and conditions
- o   The definition and analysis of risk factors associated with various terms and conditions

CRL worked with three partners (University of Illinois, Urbana-Champaign, Michigan State University, and Yale University) to explore the obligations and terms acceptable to each that would enable a long-term agreement to maintain and store selected materials. The sample contract and actual agreements with each partner institution are available online for consultation.[11]

---

[7] JSTORNEWS, No. 8, Issue 3, December 2004. http://www.jstor.org/news/2004.12/repositories.html (Accessed 10/7/2006).
[8] "UC Chosen for JSTOR Paper Repository." http://www2.library.ucla.edu/facultynews/2850.cfm (Accessed 10/7/2006).
[9] "Cooperative Collection Development." CONSORT Colleges. http://www.wooster.edu/library/oh5/cccd/CCCD_JSTORLR.html (Accessed 10/7/2006).
[10] "Distributed Print Archive Project." Center for Research Libraries. http://www.crl.edu/content.asp?l1=13&l2=19&l3=35&l4=62&l5=13. (Accessed 10/7/2006).
[11] "Distributed Print Archive Project – Terms and Conditions." Center for Research Libraries. http://www.crl.edu/content.asp?l1=4&l2=19&l3=35&l4=62&l5=19 (Accessed 10/7/2006).

This model differs from the previously existing models in that it is centrally coordinated, but not centrally collected. Participating libraries contract with CRL to serve as depositories for selected titles. The member would agree to treat its "owned material" in ways that meet CRL's guidelines. CRL would, in turn, strive to locate, obtain, and send to the member "consortium materials" to fill gaps in a member's holdings. CRL and the individual member would retain ownership of their respective titles, but the depository would manage both sets of material in similar fashion.

The model, while restrained in its initial approach, has shown some success. There have been a few similar efforts patterned on the approach, such as the ORBIS/CASCADE Alliance proposal to create a distributed print archive of two copies (one circulating; the other non-circulating) of journals from JSTOR and the American Chemical Society (ACS).[12]

Other Journal Archives:

Beyond JSTOR arrangements, there are a number of other promising arrangements regarding journals.[13]

- In 2002, the State University Library Consortium in Florida faced budget restrictions that required cancellation of print subscriptions across the consortium. Therefore, the Consortium agreed to create a policy to maintain a single print subscription for each title in an electronic aggregator package (e.g., Elsevier, Kluwer, Wiley), to which one or more of our institutions may continue to subscribe.

- The Phoenix Group of Texas academic libraries is planning the prospective cooperative acquisition of some of the same major publishers' output, to minimize or eliminate the cost of redundant print subscriptions.

- The Committee on Institutional Cooperation (CIC) Libraries are also are seeking economical ways to archive print journals. The retention of these journals in print is of high importance to faculty members because the journals sometimes differ from their electronic counterparts or contain illustrations of a quality not found in the electronic editions. Even more pressing, publishers of electronic journals, for the most part, cannot ensure the long-term availability of their products. Therefore, the CIC libraries wish to move forward plans to share the responsibility of retaining copies of print journals.

  Twelve CIC libraries will participate in the pilot project to archive a print copy of paper journals from Wiley and Kluwer/Springer. One print copy of the Wiley journals will be shipped to Indiana University and housed in its off-site facility; likewise, one print copy of the Kluwer journals will be shipped to UIUC and housed there. Each of the twelve participants will share the cost of a single paper subscription. During the pilot, CIC will track storage and processing costs and the number of ILL requests that need to be filled from the archival copy.

- The University of California Library System is undertaking a wide variety of initiatives to develop new areas of shared collections and archives and enhance current areas of cooperation. Among these initiatives are jointly negotiated electronic serial subscriptions with print counterparts intended for shared archiving, and plans to create a shared archive of government documents.

---

[12] Orbis Cascade Alliance. Collection Development and Management Committee. "Proposal: Pilot Project for a Distributed Print Repository." September 2005. http://www.orbiscascade.org/staffhome/CDMC_distributed_print_archive_12-9_final.doc . (Accessed 10/7/2006).
[13] Text selected from http://www.crl.edu/content.asp?l1=13&l2=19&l3=35&l4=64

Non-Journal Activities

Serial collections aside, there is interest in preservation of other print collections. Many of these discussions are still in preliminary discussions, and the proposed models in flux.

Government Documents

The Government Printing Office (GPO) of the United States is working with the federal depository library community on several new initiatives involving access to and preservation of print and digital federal government documents.

The Federal Depository Library Program (FDLP) was established in 1895 to provide US citizens with "no-fee access" to federal government information. The distribution of materials is coordinated through the GPO.[14]

In laying out a strategic vision and action plan for the GPO in the 21st Century[15], the GPO described the challenges facing the government and FDL libraries. In the past decade, significant changes in how information is distributed and accessed have impacted the role depository libraries. The increasing popularity of the World Wide Web as the preferred medium through which the public seeks access to information as well as the shift to electronic publishing of government documents have changed the nature of the federal depository system. The GPO is currently developing its Future Digital System (FDsys) which will store and serve all digital publications entering into the FDLP, effectively making available on the Web the collection of materials traditionally accessed through FDLs.

The GPO envisions an ongoing need for depository libraries, both serving an *access need* for constituents requiring print versions, and a *preservation need* as a backup to ensure long-term access to US government information. The GPO has encouraged the movement toward **shared repositories**, or shared housing agreements, that allow libraries to eliminate some of the redundancy among their collections. The GPO contracted with CRL to develop a *Decision Framework for Federal Document Repositories* that can be used to evaluate the level of assurance provided by such repositories, based on their physical characteristics, resources, governance and other factors.

To provide added assurance, the GPO has developed a draft plan for a **collection of last resort**; a secure "dark archive" that will become, over time, a comprehensive collection of tangible and electronic titles that will backstop the regional library collections and future shared repositories. Through this collection, GPO will support permanent public access to U.S. Government publications in all formats.

CRL is also working to identify the **risk factors involved with long term retention**. The Superintendent of Documents of the U.S. GPO recently asked CRL to incorporate these issues in their thinking about a possible repository system for legacy government documents. A discussion draft of the report is available from the link http://www.access.gpo.gov/su_docs/fdlp/pubs/decisionmatrix.pdf .

---

[14] "Federal Depository Library Program." http://www.arl.org/info/frn/gov/fdlp/index.html (Accessed 10/7/2006).
[15] Government Printing Office. Depository Library Council. "The Federal Government Information Environment of the 21st Century: Towards a Vision Statement and Plan of Action for Federal Depository Libraries." September 2005. http://www.access.gpo.gov/su_docs/fdlp/pubs/dlc_vision_09_02_2005.pdf (Accessed 10/7/2006).

Monographs

ASERL Virtual storage[16]
The Association of Southeastern Research Libraries (ASERL) based in Atlanta, Georgia, has embarked on a pilot study of creating a "Virtual Storage" arrangement with its members. Since the early 1990's, the Southeastern Library Network (SOLINET) has discussed building a regional storage facility to be shared by its 2300 members. Many libraries expressed interest, but several factors impeded implementation: aside from the cost concerns, state-sponsored libraries worried that sending their material out of state would violate their charters, and others were concerned that deaccessioning or discarding multiple copies would negatively impact their standings in comparative statistics and national ranking (ARL volume counts).

ASERL Proposed in 1999 that a "virtual" storage concept may be more appropriate than building a shared facility. ASERL operates a shared catalog and delivery system. ASERL began working on a system, beginning with monographs (as they were more easily compared bibliographically than serials). By identifying unique materials in existing remote storage facilities and tagging them in OCLC, members could be reasonably assured that these materials would be preserved in quality environments and not be deaccessioned. Such a system would allow libraries to consider discarding items that are known to be held in the national/regional preservation collections, thereby avoiding sending yet another copy of the item into storage. Each library could arrive at its own comfort level about how many copies need to be held in the national/regional system.

In a proof of concept study in 2004, nine participating libraries compared their remote storage holdings. The findings showed that 80% of all collections (2.3 million monographs) were held by only one library. Only 7% of the collection was held by three or more of the participating libraries.

OCLC played a key role in the study, providing a comparison of data by participating libraries and as an indicator that any particular copy is the "preservation copy."

University of California System
As previously discussed, UC's shared repository has investigating a range of formats and materials. In addition to JSTOR journals, and journals and other publications of the Association for Computing Machinery (ACM), the California libraries are taking initial steps to pursue centralized deposits of the shared microfilm subscriptions to the Short-Title Catalogue (parts I and II) as a shared analog counterpart to Early English Books Online. Campuses are sharing the cost of this subscription. Discussions regarding other monographic collections (either digitized or not) are still underway. This includes a proposal to share monographs through a Small Press Distribution Approval Plan Scenario.

Conclusions:

The variety of print archive activities should provide models for institutions and clusters of organizations interested in preserving resources for the future. A multiplicity of studies and assessments are available online for consideration. The Center for Research Libraries provides a gateway to a number of efforts underway.

A large majority of efforts are devoted to preserving material being made available online through commercial or non-commercial providers. Print archives are tied inextricably to digital efforts and the ongoing concern over the longevity of electronic information. Print archives are a necessary, if perhaps transitory, component in the evolution of "trusted digital repositories."

---

[16] "ASERL's Virtual Storage/Preservation Concept," presented at ACRL 12th National Conference, April 7–10, 2005. http://www.ala.org/ala/acrl/acrlevents/burger-etal05.pdf. (accessed 10/7/2006).

The University of California system has engaged in perhaps the widest variety of activities and has engaged in numerous assessments of their efforts. In their investigations, UC has pointed to a number of advantages. In the planning framework for their shared print collections, they state:

> These collections realize the priorities and opportunities for cost avoidance that emerge in the development, system-wide, of our shared digital collections. The Elsevier pilot has clearly demonstrated significant cost-savings that can be realized…. Typically, they involve the acquisition for the system of a single set of print journals for all of the titles to which the UC libraries also subscribe electronically. The existence of the so called "print archive" allows campuses that wish to do so to cancel local print subscriptions in favor of electronic only ones without denying faculty who need it access to the print journals.[17]

At the same time, the university points out a number of challenges that will need to be addressed in the long term. For instance, cost savings may be reduced as journal publishers increasingly move to e-only production formats. As digital versions become the "copy of record," and preservation systems of digital resources are put in place, the role of print collections should revisited.

Costs of "dark archives" may prove untenable in the long term. While projects limited in scope are reasonably affordable to implement, larger shared collections will take up increasing amount of costs and staff time. Most institutions have shown reluctant to contribute long-term funding for projects that do not provide access to materials. For this reason, digital archives that store but do not hold the promise of long-term access to materials have proven similarly difficult to succeed.

A broader concern is the cost of preserving the materials for the "greater good." Should one institution shoulder the burden of preserving materials that will clearly benefit other libraries that are not committing similar resources to other materials? A system of common expectations and proportionate commitments would better justify the large investment required to supporting such an initiative in the long term.

As the "Evidence in Hand" report points out, "American scholarship has historically relied on a decentralized network of libraries, archives, historical societies, and museums to collect, preserve, and make accessible the intellectual and cultural resources that form the basis of academic inquiry. This decentralized and largely uncoordinated approach is unlikely to change, despite the resulting losses of artifacts that have occurred."[18] While the US system allows for a certain level of security through the distribution of multiple copies of materials across the country, it does not create a facile system for the designation of institutions of record that could be charged with preserving classes of material.

In Germany, the system of nationally coordinated *Sammelschwerpunkte* (Areas of Collection Emphasis) distributed throughout the Federal Republic might provide such assurance that the broad panoply of institutions could be responsible for the long-term maintenance of classes of material aligned within the designated specialized disciplines. It is notable that priorities for long-term preservation of information resources in their original form is absent from the Deutsche Forschungsgemeinschaft's "Funding Priorities Through 2015" position paper.[19] This topic, of concern to libraries in Germany and among the international community, should be discussed and brought to the attention of planners and funding organizations at the national level.

---

[17] "Developing a Planning Framework for UC Libraries Shared Print Collections." April 14, 2005. http://www.slp.ucop.edu/programs/sharedprint/PlanningFrameworkv1-8.pdf (Accessed 10/7/2006).
[18] "The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections." p. 15.
[19] Deutsche Forschungsgeneinschaft. "Scientific Library Services and Information Systems: Funding Priorities Through 2015." http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_funding_priorities_2015_en.pdf. (Accessed 10/7/2006).