



The Open Past Initiative: A Discussion Paper

DRAFT

Preface

The purpose of this paper is to initiate a discussion among SPARC, librarians, publishers, and other interested parties of a potential initiative to digitize and disseminate the back-runs of scholarly journals. SPARC seeks to obtain feedback on the concept described here; to determine whether the fundamental approach has sufficient merit to warrant further planning and development; and to elicit preliminary expressions of interest in participation.

Situation Assessment

The demand for electronic access to scholarly journal research articles continues to mount across all disciplines. Yet, of the thousands of academic journals now publishing electronically, only the largest and richest can afford to convert their entire run of back issues into digital format. This leaves the historic backfiles of thousands of academic journals locked in print, with access to each limited physically to a relatively small number of libraries.

This situation benefits no one. Library investments in materials go underutilized, publisher assets lie unproductive, researcher access remains limited, and author research impact atrophies. Providing free electronic access to these valuable research resources would therefore benefit all the stakeholders in the journal publishing process:

- Researchers would gain increased access and greater functionality;
- Authors would experience increased impact and visibility for their research;
- Publishers would enjoy greater use of their content, which would build journal prestige, attract prospective authors, and generate new subscriptions to current journals; and
- Librarians could enhance patron service, while lowering storage, retrieval, and interlibrary loan costs.

While JSTOR provides fee-based access to digital backfiles of selected journals, it can only address a fraction of the retrospective print journal content. The potential SPARC program described aims to provide a practical cooperative framework through which academic research libraries and scholarly journal publishers can expand the body of digital content available to better serve their own constituencies and the broader interests of scholarly communication.

Proposed Response

SPARC proposes launching a cooperative program—tentatively titled the “Open Past Initiative”—that would apply the digital conversion and content management expertise of academic libraries to unlock the now isolated print assets of nonprofit and other independent

journal publishers. Participating publishers likely would be insufficiently capitalized or otherwise ill equipped to undertake conversion on their own. They would contribute non-exclusive electronic license rights to their backfile content, while participating libraries would contribute retrospective conversion resources and long-term web-based open access. All researchers worldwide would have free online access to the valuable digital content thus made available.

Program Description

The Open Past Initiative would provide an exchange forum that would help match printed journal backfiles requiring retrospective digital conversion with academic libraries willing to support the conversion and online availability of the journal's retrospective content.

SPARC's Role & Program Structure

As the program's catalyst, SPARC would define and promote the Open Past Initiative; create an online matching clearinghouse for potential program participants; provide a model library-publisher agreement; supply business advice and guidance in brokering arrangements and negotiating agreement terms; facilitate community agreement on use tracking, interoperability, and other technical standards; and actively market and publicize the benefits of the program to both publishers and libraries.

Should institutional participants indicate sufficient interest, SPARC would negotiate preferential rates with digital conversion vendors. This would take advantage of any economies of scale the initiative might afford, and also allow participants to project the potential cost of the digital conversion prior to committing to a publisher.

The Open Past Initiative would be a cooperative program and its activities would be geared to coordinating and facilitating the collaborative activities of libraries and journal publishers rather than with prescribing and monitoring technical standards.

The program would be instituted in several phases. The initial phase would match several institutions with appropriate journals on a pilot basis to demonstrate the concept and to provide the practical experience necessary to develop the supporting program infrastructure.

Subsequent phases could:

- convene a content selection group to identify high-value retrospective journal content, and actively recruit publisher participants and match them with suitable institutional partners;
- identify institutional partners willing to serve as centralized repositories for the converted content (Cornell and University of Michigan have already expressed preliminary interest);
- develop a sponsorship program—corporate and/or philanthropic—to help libraries underwrite part of the cost of converting and hosting/maintaining the journal content; and
- create other scale-based support programs.

Library Role

Participating institutions would commit to underwriting the one-time expense of digitally converting a journal's print backfile and to the long-term storage, maintenance, and hosting of the content online. Ideally, the conversion would cover the complete print back run of a journal, but practical considerations—the availability and/or condition of the print, copyright control, and other issues—might lead the institution and the publisher to mutually agree on a subset of the backfile. In some instances, a cooperating institution and publisher might agree to add born-digital journal content to the open-access backfile on a prospective rolling window basis.

Libraries would “adopt” journals having content that supports the institution’s academic teaching and research programs. From a practical perspective, this would frequently reflect an especially strong faculty affiliation with, and influence on, a particular society publisher. Selecting journal content that aligns well with its academic programs would help an institution justify the initial expense and ongoing obligation of maintaining the content availability online.

Publisher Role

The journal publisher would grant the sponsoring institution a non-exclusive and perpetual open-access license to digitize, maintain, and distribute the content. Other provisions of the distribution license would typically include the responsibilities of each party, agreed upon technical specifications, the scope of coverage (including prospective years, if relevant), conversion and online availability schedule, and other customary terms and conditions.

The publisher would retain the right to offer the digital content in services of its own, including gated proprietary services, without impact on the open access distribution rights of the institution.

Besides the program-specified conversion and maintenance terms, participants could agree to collaboratively develop additional service components—for example, enhanced access tools and cross reference linking—that one or both parties could make available on a fee basis, depending on specifically negotiated terms. Further, multiple institutions and publishers could collaborate to create multi-publisher content sets and/or to undertake a conversion project beyond the resources of a single institution.

Scope & Cost Issues

Conversion Costs

Table 1 illustrates the potential digital conversion costs for a 25-year print journal backfile. While actual costs would vary based on a host of variables, Table 1 provides a conservative estimate of the potential conversion costs to a participating library. The actual digital conversion costs per journal would represent a relatively small proportion of the overall costs. Much of the cost would be in the form of either existing or incremental staff time.

Table 1: Average Conversion Costs

Backfile Pages/Year ¹	Backfile Years	Conversion Costs/Page ²	Conversion Costs/Journal	Processing Costs/Page ³	Processing Costs/Journal	Total Costs/Journal
1,813	25	\$ 0.35	\$ 15,864	\$ 0.25	\$ 11,331	\$ 27,195

1) Average annual article page estimate for society published journal. Source: Tenopir & King (2000), Table 58.

2) Includes digital scanning (without debinding), OCR, and tagging; quality control; and other processing costs.

3) Includes metadata creation and other staff pre- and post-conversion processing and management costs; figured at 43% of total costs. See Chapman and Kenney (1996), Table 4.

Storage & Hosting Costs

Hosting costs would represent a modest incremental expense for institutions with existing capacity. Additionally, the initiative would encourage institutions to cooperate in establishing mirror sites to ensure redundant storage for the material. As Table 2 indicates, the average storage space requirements for a 25-year journal backfile run would be relatively modest in the context of a university computing environment.

Table 2: Average Storage Space Requirements

Backfile Pages/Year ¹	Backfile Years	Storage Size KB/Page ²	Backfile Size GB/Journal ³	Cost per GB ⁴	Backfile Cost per Journal
1,813	25	130	7.02	\$ 15.00	\$ 105.36

1) Average annual article page estimate for society published journal. Source: Tenopir & King (2000), Table 58.

2) File size assumes a 600 dpi bitonal with a descreening filter, producing a lossless compressed file of 127KB, plus 3KB for OCR'd text.

3) Storage requirement includes a 25% space redundancy allowance.

4) Based on OCLC archival storage costs. Local storage costs would be lower.

Digital Preservation Costs

Long-term digital preservation costs are impossible to determine. However, complying with best practices for digital conversion and metadata creation would help ensure that the digital content converted via the initiative could be cost-effectively maintained on a long-term basis. Further, participation in LOCKSS would help ensure the ongoing availability of the converted material. LOCKSS requirements and membership could be negotiated by SPARC for all participating institutions.

Technical Issues

Conversion & Preservation Standards

Guidelines for technical specifications, suggested by a library-publisher advisory group, would address such matters as digital format(s), DOI registrations, citation linking, and related issues.

Given the rapidly evolving technical environment for digital conversion and preservation, the initiative would not prescribe exact technical specifications. Rather, it would only suggest guidelines conducive to the long-term availability and interoperability of the digital resources, and point to existing technical standards and best practices promulgated by expert organizations.¹

The digital format(s) selected by any matched pair of program participants would depend on a variety of factors, including existing conversion and hosting practices at the institution and/or the digital format of a journal's prospective material. As long as the institution agrees to ensure the long-term digital preservation of the material and the publisher provides the requisite open-access rights, the purpose of the initiative would be fulfilled. Baseline metadata and interoperability standards would then suffice for discovery of the material in a distributed repository network.

The program would encourage implementation of a standard cross-reference protocol that would link articles digitized by the program.

Interoperability Protocols

The program would encourage compliance with the Open Archives Initiative Metadata Harvesting Protocols, and other standards as may prove appropriate, to facilitate the interoperability of distributed digital repositories containing the converted content.

Online Usage Metrics

The program would need to ensure that adequate provisions are in place for tracking and reporting usage of the converted material. This usage data would demonstrate the success and value of the program to both libraries and to journal publishers and their authors.

¹ These might include the Digital Library Federation, the Northeast Document Conservation Center, and Cornell University.

Potential Issues & Responses

As virtually all academic institutions worldwide would benefit from the creation of the openly accessible backfiles, the free rider issue would require that the program demonstrate compelling reasons for individual institutions to participate, rather than to wait for others to act. Some libraries would participate in the program given their affinity with the program's goals. However, for the program to succeed on a large scale would require that institutions perceive the benefits of providing the content to their own institutions as commensurate with the financial investment being made.

Program Economic Benefits

For many institutions, the increased use, broader accessibility, and superior functionality afforded by a digital version of the content would provide the impetus for their program participation. Still, this retrospective digitization program would also deliver quantifiable economic benefits to participating libraries and others. Some of these benefits are described below.

ILL Cost Savings

Over time, as the body of retrospective journal content made available via the Open Past Initiative increases, interlibrary loan borrowing and lending costs could decrease for all institutions. Assuming that journals represented by the program represent, on average, one-half of one percent of ILL transactions, each library could save approximately \$3,454 per year (see Table 3). While savings might be modest for any given institution, overall this could represent an annual savings of over \$425,000 per year for ARL institutions alone.

Table 3: Potential ILL Savings (per institution)

Transaction Category	Total Transactions/Yr ¹	Pct. of Total Transactions	Avg. Cost/Transaction ¹	Potential Savings
Borrower	21,323	0.5%	\$ 14.98	\$ 1,597
Lender	29,108	0.5%	\$ 12.76	\$ 1,857
All	50,431	0.5%		\$ 3,454

1) ARL figures.

Freed Shelf Space

Similarly, libraries could realize savings on shelf space. As most participating libraries would already own the print runs for any journals they sponsor, they could conceivably realize savings by shifting the corresponding print volumes from on-campus stacks to an offsite storage facility. As Table 4 indicates, every 100 volumes shifted from on campus stacks to offsite storage subsequent to digitization would result in an approximate savings of \$1,200 per institution.²

Table 4: Potential Print Storage Savings (per 100 volumes)

Issues per Volume	Backfile Years	Total Volumes	On-campus Storage/Vol ¹	Offsite Storage/Vol ¹	Potential Savings
4	25	100	\$ 16.00	\$ 4.00	\$ 1,200

1) See Kevin Guthrie. "Archiving in the Digital Age." *Educause Review* (Nov/Dec 2001).

² While the library would not realize an actual budget impact given academic accounting practices (which typically separate capital and operational expenses), the space savings themselves would be real.

Next Steps

If this discussion paper generates positive feedback from SPARC members, SPARC will:

- Modify the program concept based on SPARC member feedback and circulate the revised concept description.
- Conduct two surveys:
 - Survey SPARC and SPARC Europe members to assess their willingness in principle to sponsor the conversion and hosting of journal content as envisioned by the initiative; and
 - Survey a representative group of society journal publishers to assess their willingness in principle to provide open-access distribution licenses and participate in the Open Past Initiative.

If these surveys indicate sufficient interest on the part of the universe of potential program participants, SPARC will:

- Develop a formal program proposal.

If the formal proposal is approved by the SPARC steering committee, SPARC will:

- Prepare a detailed initiative plan—including a fuller description of the support programs needed, implementation schedules, and budgets—with the objective of launching the initiative by summer 2004; and
- Project detailed program costs and identify potential funding models, including a prospectus to support potential sponsorship underwriting.

To Comment

SPARC encourages your comments and questions regarding the program outlined in this discussion paper. **Please send your comments to: Rick Johnson at rick@arl.org.**



SPARC
21 Dupont Circle, NW, Suite 800
Washington, DC 20036 USA
Tel: 202-296-2296 Fax: 202-872-0884

SPARC, the Scholarly Publishing and Academic Resources Coalition, is an initiative of the Association of Research Libraries.