Ithaka Boston Forum

Andrée J. Rathemacher
University of Rhode Island, andree@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/lib_ts_pubs

Part of the Library and Information Science Commons

Terms of Use
All rights reserved under copyright.

Citation/Publisher Attribution
Ithaka Boston Forum

Andrée J. Rathemacher

On December 6, 2011, Ithaka S+R hosted an Ithaka Forum in Boston. Presenters from Ithaka S+R updated attendees on new initiatives at JSTOR, forthcoming Books at JSTOR, the JSTOR Current Scholarship Program, and long-term preservation of electronic content through Portico.

1. Extending JSTOR’s Mission

Bruce Heterick (vice president, Outreach and Participation Services, Portico and JSTOR) spoke on new initiatives at JSTOR, the first of which was the launch of Early Journal Content in early September 2011. More than half a million journal articles from over 200 journals published before 1923 in the United States or before 1870 elsewhere are now freely available on the JSTOR platform. No institutional affiliation or registration is required to access this content, which is being heavily used.

Another initiative is the Register and Read Program. Heterick explained that indexing of JSTOR content by Google and Google Scholar resulted in close to 200 million attempts to access the JSTOR database last year. Unfortunately, only 20 percent of users could be authenticated and provided with access, resulting in 180 million bad user experiences. The Register and Read program is an attempt to get JSTOR content to people who are unaffiliated scholars. In this pilot program, unaffiliated users who create a MyJSTOR account are provided with free, read-only
access to up to three articles at a time for a specified number of days (currently fourteen in the beta test). At or before the end of this period, the user may purchase and download the article for future reference. The hope is that this “freemium” model will encourage users to return to the JSTOR platform and will provide them with a more positive impression of the journals and publishers whose content they wish to access. Though publisher participation in the pilot is limited (only about fifty journals are currently included, and only archival articles are for sale), JSTOR is encouraging these publishers to experiment with lower per-article pricing to increase the percentage of readers who decide to purchase an article.

In another current program, JSTOR is allowing the alumni of nineteen subscriber institutions to access the JSTOR database. During the pilot phase, these institutions are paying an additional ten percent on top of their annual access fees as JSTOR and the institutions evaluate usage, funding methodologies, and successful marketing tactics. Additional institutions have signed up to participate in a second phase of the pilot during 2012. After the pilot is over, JSTOR will offer alumni access to additional institutions at a price no higher than the costs of administering the program.

Heterick’s favorite JSTOR initiative is called “Institutional Finder (Proxy Re-direct).” Heterick pointed out that a significant number of researchers affiliated with JSTOR-subscribing institutions discover JSTOR content through Google. Many are not logged in through their institutions’ networks and are therefore denied access. The Institutional Finder (Proxy Re-direct) presents unauthorized users with a message informing them that full text might be available if
they are affiliated with a participating library or publisher and prompts them to log in with their MyJSTOR Account or institutional credentials. Once a user identifies her institution, she may then click on a link to her institution’s login page provided by JSTOR. After she logs in, she will be returned to the requested article. Institutional Finder (Proxy Re-direct) is functioning now for over twenty-five institutions and works with most authentication systems.

Heterick spent the remainder of his time discussing the broad issue of content discovery in general and how various discovery services are using JSTOR metadata. JSTOR’s goal, in addition to preservation, is to extend access to JSTOR content to as broad an audience as possible. To this end, JSTOR has been providing their metadata to discovery platforms including Serials Solution’s Summon, EBSCO Discovery Service, Ex Libris’ Primo, and OCLC WorldCat. Now JSTOR is looking at how their metadata is being used by these services and if it is being represented properly.

Heterick lamented the lack of transparency in how these services determine relevancy and ranking and the impact this has on the publishers and providers, like JSTOR, that provide them with metadata. JSTOR has noticed that use of JSTOR at an institution tends to decline once the institution has implemented a discovery service. Heterick explained that a search for “global warming” in JSTOR produces tens of thousands of hits, while the same search on certain discovery platforms produces fewer than 2000 results from the JSTOR database. Heterick worries that the vendors of some of these discovery platforms might be gaming the system by using indexing algorithms that optimize their own content. Since librarians make subscription
decisions based on usage, this is an important issue for JSTOR and their partner publishers. Heterick complained that discovery service relevancy rankings are like a “secret sauce” whose recipe is closely guarded. He hopes that the National Information Standards Organization’s (NISO) Open Discovery Initiative will address this lack of transparency through the development of standards and best practices. In the next three to five months, JSTOR, though Ithaka S+R, will also be studying this issue with the goal of being able to provide guidance to libraries about which administrative settings in their discovery products will affect the ranking of JSTOR content.

Following from this discussion of discovery platforms, Heterick transitioned into some reflections on the declining importance of the library and library platforms in general in discovery. A 2009 report by the University of Minnesota Libraries found that users are successfully discovering relevant information resources through non-library systems such as general Web searches and social networking applications (Hanson, et al., 2009). Similarly, in Ithaka’s 2009 Faculty Survey, fewer faculty reported starting their research in online library catalogs than in previous years, and more faculty reported that general-purpose search engines or specific electronic research resources were their starting points (Schonfeld, 2010). Backing up these findings are recent data on where JSTOR sessions originated between January and October 2011. Fifty-six percent of sessions originated in Google; 19 percent in JSTOR; 10 percent through linking partners; 8 percent through other sites including CrossRef, Wikipedia, ISI, and facebook; and only 7 percent from libraries.
The University of Minnesota Libraries report recommended the principles that discovery be “organized around users rather than collections or systems,” and that to be good stewards of collections means “participating in cooperative ventures that provide broad access to… collections” (Hanson, et al., 2009). With this in mind, Heterick advocated a “push, don’t pull” strategy in which “the goal should not be about trying to bring the researcher back to the library; the goal should be how do we better bring the local library resources to the researcher from wherever they happen to begin their research (including Google).” For Heterick, this not only raises the question of how to best make JSTOR metadata available on other search platforms including library discovery services but, if library users are starting their research at JSTOR, how to make them aware of a library’s other licensed content available elsewhere.

To this end, JSTOR has developed the Local Discovery Integration (LDI) Pilot, which will allow researchers searching JSTOR to link out to their libraries’ local discovery services. Twelve libraries that have implemented Summon, Ex Libris Primo, EBSCO Discovery Service, or WorldCat are participating in the pilot, which started in April 2011. After a user completes a search in JSTOR, a link appears on the search results page that allows her to repeat the search in her library’s discovery service in a new browser window. The link also appears on the Advanced Search screen, on the Article View page, on the screen that appears if no results are found in JSTOR, and in a pop-up box on the third page of JSTOR results. So far, most clicks on the LDI links have come from the no results page followed by the initial search results page and the third search results page, though only about 1 percent of users have been clicking on the links at all. JSTOR is studying not only which icons and languages work best for users, but what happens after the discovery service window opens. Heterick pointed out that not many resource
providers would be willing to drive traffic away from their own sites in this way, but if they find that it works for users, it is something that they will consider incorporating into the JSTOR interface.

2. Unpacking Books at JSTOR

In the second morning session, Frank Smith (director, Books at JSTOR) introduced Books at JSTOR, an e-book initiative that will launch in summer 2012. Frank introduced himself as new to JSTOR, having spent most of his career at Cambridge University Press, where he lead the team that built Cambridge Books Online. He commented that publishers are “making this up as they go along,” and that “the only rules are that there are no rules.” Smith wants to see scholarly books continue to be published, and he doesn’t care whether they are published in print or digital format.

Smith posed the rhetorical question, “Why books on JSTOR?” One reason is that JSTOR has a very high degree of awareness and reputation in the scholarly world. Ithaka S+R Faculty Survey respondents reported using JSTOR more frequently than Google Books, Project Muse, EBSCO, and other resources and gave JSTOR the highest rating in terms of “overall image.” In addition, interest in e-books is growing. A survey of acquisitions librarians at JSTOR participating libraries revealed that e-books were the most desired content addition to the JSTOR platform, and 72 percent of faculty in Ithaka S+R’s Faculty Survey 2009 indicated that digital versions of scholarly books on the JSTOR platform would make JSTOR more valuable to them. Finally, the journal content on JSTOR would integrate well with books, as there are currently an estimated
600,000 citations to books and nearly 2,000,000 book reviews in the JSTOR database. Thus, the inclusion of books in JSTOR will enhance the researcher experience by incorporating cited materials on the same platform.

JSTOR has partnered with about twenty-seven publishers including some of the largest university presses (e.g., Harvard University Press, Columbia University Press, University of Minnesota Press). In addition, Books at JSTOR will include books from well known public policy presses including the Brookings Institution, the RAND Corporation, the Russell Sage Foundation, and United Nations Publications. JSTOR is working with Boydell & Brewer and Edinburgh University Press in the United Kingdom, McGill-Queen’s University Press and University of Toronto Press in Canada, and CSIRO and the University of Aukland Press in Australia. Smith said that this list is only a starting point, but it represents a strong start with publishers of books that are at the top of the list of what scholars and librarians prioritize. He noted that fifteen of the top twenty-five most-cited university presses are in Books at JSTOR as are four of the ten most-cited publishers.

Searching for books on the JSTOR platform will be integrated with searching for articles and other content. Book content will be delivered at the chapter level, which has become standard across e-book publishers. JSTOR’s hypothesis is that this “soup” of blended scholarly content will be valuable to researchers; they will have to determine if this proves true. The integrated search results will allow for faceting by content type (book chapters, journal articles, reviews, and primary sources), by broad topic area, and by date and decade. Dragging the mouse over a
“Quick View” icon on the results page will open a window revealing the book’s metadata along with brief chapter and book summaries. The Chapter Preview will present the first few pages of each chapter and a chapter summary to help researchers decide if they want to read further, and links to book reviews appearing elsewhere in JSTOR will be provided. Deep linking will connect journal and e-book content throughout the database.

When Books at JSTOR launches in July 2012, books will be available through a purchase model only as opposed to a subscription model. Books will be able to be purchased in any quantity as firm orders or through a demand-driven model. Though books will be available individually, JSTOR anticipates working with libraries to create subject and publisher collections that may vary from institution to institution. Libraries may purchase books through YBP Library Services or directly from JSTOR. The price for each title will be set by the publisher and will be tiered by JSTOR Class with different prices for frontlist and backlist titles and volume discounts available. All titles will be preserved in Portico.

There will be two access models for Books at JSTOR: unlimited and limited. The unlimited model will allow for an unlimited number of simultaneous users and unlimited online reading. Chapters may be downloaded according to license terms and conditions; they will be watermarked and IP-tagged but will not be subject to DRM or printing restrictions. Not all books will be available in the unlimited model; publishers will decide which are. Though publishers will set their own pricing, JSTOR expects that the maximum price for an unlimited-
use book will be about thirty to forty percent above list price. Since pricing will be tiered by JSTOR Class, smaller institutions will pay less.

Books available through the limited model will be viewable by only one user at a time. All books will be available in the limited model; some books will only be available in the limited model. Single users may view limited model books online for an unlimited amount of time. Chapter downloads will be limited to 30 per year, though libraries will be able to purchase additional downloads and up to three additional “copies” of the book. Downloaded chapters from limited model books will have DRM to prevent sharing and some restrictions on printing. The maximum price for limited-use titles will be the highest print price, with smaller institutions in lower tiers paying less.

Smith explained that the limited model was necessary because most publishers receive a majority of their revenue from a small number of titles that are used heavily as assigned reading in classes. If these books were available under the unlimited model, publishers might lose this revenue stream, which would threaten their financial stability. Other online book platforms, for example those from Oxford University Press and Cambridge University Press, offer all their e-books in the unlimited use model, but 30 to 40 percent of their titles are not available online for this reason. JSTOR’s concern was that they have as many of the best books as possible in Books at JSTOR, so they needed to find a way to balance availability with the need to protect publishers’ revenue streams. The percentage of books available through the unlimited model
will vary widely by publisher based on what portion of their title list is geared primarily toward researchers versus researchers and students.

As an aside, Smith noted that the level of risk to publishers of offering unlimited access to scholarly books in library settings is not known. There are no good data to support how much users will read online, download, and print. Currently, a great deal of scanning and online sharing takes place illegally, and students prefer printed books. In response to one attendee’s contention that the limited use model was not helpful for libraries, Smith agreed, saying that JSTOR would like to see publishers move toward allowing more access. Heterick added that JSTOR’s goal is to launch Books at JSTOR with as many books as possible. If JSTOR had insisted on the unlimited model for all books, publishers were saying that they would withhold about 60 percent of their corpus. JSTOR’s goal is to build the Books at JSTOR platform first and then fix it later. Heterick stated that JSTOR and Portico have always started in uncomfortable places. At first, most journal publishers with content in JSTOR insisted on a moving wall of seven to ten years; now the moving wall has decreased to three years on average as publishers have realized that including their archival content in JSTOR does not cannibalize current subscriptions. Likewise, when Portico began most publishers would not allow Portico to provide post-cancellation access, but now more than 90 percent do. Heterick believes the same thing will happen with Books at JSTOR—once publishers become comfortable with the platform, they will reduce restrictions on access.
An attendee countered that because of these limitations, libraries’ interest in purchasing e-books on the JSTOR platform will be diminished. Smith responded that publishers will learn from libraries and adapt. He explained that book publishers do not talk to libraries and know very little about what they do. Book publishers work with multiple sales channels, and institutional sales are only one small component, which they do not think about very much. Traditionally, wholesalers stood between publishers and their customers. E-books will change this as publishers deal directly with libraries through JSTOR, which will be flooding them with usage and sales data and sharing libraries’ concerns. Another attendee complained that students will be confused about why they can access and download content from some books but not others; there needs to be a way to easily identify limited access books so librarians can explain the difference to library users. Smith answered that there will be a clear message at the book and chapter levels when a title is limited-use. Heterick added that through an administrative module, librarians will be able to purchase additional downloads, set notifications when downloads for a title are about to run out, and turn on or off a link that will allow users to individually purchase a chapter or a book.

Smith continued, stating that all e-books on the JSTOR platform will have quality metadata including digital object identifiers at the chapter level and MARC records, which will be “imperfect.” A number of additional questions were raised by attendees. One concerned the degree of overlap with ebrary. Heterick responded that there will be overlap, but that JSTOR would work with libraries to avoid duplicate purchases and investigate possibilities for providing access through JSTOR to books purchased on other platforms. Further questioning revealed that purchasing an e-book on JSTOR will be completely independent from the purchase of the book.
in print; if a library buys a book directly from JSTOR, JSTOR can report the purchase to YBP Library Services so that they know that the library owns the title. The availability of backlist titles on Books at JSTOR will vary tremendously by publisher based on how many older titles they have digitized. There will be no support for print-on-demand, at least not initially. Interlibrary loan of chapters will not be allowed, and JSTOR has plans to develop a patron-driven, short-term loan option in the future, in part to address concerns about interlibrary loan.

3. Thinking in Threes: The Impact of the Current Scholarship Program on Publishers, Librarians, and Users

Mary Rose Muccie (director, Current Scholarship Program) spoke next about JSTOR’s Current Scholarship Program (CSP). She outlined the challenges and motivating factors that prompted JSTOR to create the CSP; the program’s objectives and details about how it works; the program’s value propositions for libraries, publishers, and users; and the impact of the CSP on journal usage one year after initiation.

JSTOR’s CSP, which publishes current journal issues from selected publishers alongside archival content on the JSTOR platform, was formed in response to the budgetary challenges facing libraries and independent publishers. As “big deals” with major publishers absorbed larger percentages of library budgets, subscriptions to titles from independent publishers were at
risk. In addition, libraries faced increased transaction costs in negotiating multiple licenses with smaller publishers. On the publisher side, costs were increasing at the same time that revenues were threatened by declining library expenditures.

JSTOR had long been aware through surveys of librarians and faculty that demand existed for current content on the JSTOR platform. When JSTOR started looking into adding current content a few years ago, they sought to leverage the scale of JSTOR in a way that would be transformational for publishers, librarians, and users—as opposed to being “just another platform.” The CSP sprang from a desire to craft a sustainable publishing model that embodied academic values and ensured long-term diversity in scholarly publishing. It also had the objective of enhancing users’ research efficiency through seamless access to quality current and archival content.

The Current Scholarship Program was launched on January 1, 2011, with 175 titles from nineteen university presses, scholarly societies, and independent publishers. In 2012, there will be 215 titles from a total of thirty publishers. In addition to current issues, archival content is available for all CSP titles, and all titles are preserved in Portico.

Through the CSP, JSTOR is able to offer a number of benefits to journal publishers. JSTOR provides a robust technological platform for hosting journal content with tools and features valued by libraries and researchers. With the launch of the CSP, JSTOR upgraded its platform to
support multimedia including audio, video, and zoom-able images, thereby expanding the capabilities of published research. JSTOR offers value to publishers through ordering and access management, customer and end-user support, and global institutional sales and marketing support that offers exposure to market segments that can be difficult for smaller publishers to reach. The JSTOR platform provides access management tools for individual and member subscriptions in addition to institutional subscriptions. Though CSP publishers set their own subscription prices, the JSTOR model offers publishers the opportunity to implement tiered pricing, which can increase revenues.

For libraries, the Current Scholarship Program integrates JSTOR archive collections with current journal content on a single platform. JSTOR offers flexible subscription options; titles may be subscribed to individually or in packages and may be ordered directly from JSTOR or through a subscription agent. The CSP saves libraries time in that only one license and point of purchase is needed for multiple publishers and titles. In addition, library users benefit from the ability to access current content from within the familiar and trusted JSTOR interface as well as from the addition of supplementary material and multimedia.

Muccie highlighted how, as expected, the integration of CSP content on the JSTOR platform has driven increased use of these titles. JSTOR compared usage data for journals from two publishers at eleven higher education institutions from January through June of 2010 with the same period in 2011. In 2010, the institutions accessed subscribed journals on the publishers’ previous platforms while in 2011, they accessed the exact same journals on the JSTOR platform.
For one publisher’s titles, the average number of successful full-text article requests increased by 94 percent, and for the second publisher’s titles they increased by 226 percent. When JSTOR compared usage data for twelve individual titles in different disciplines, every title except for three showed a significant percentage increase in use between 2010 and 2011. Muccie remarked that one publication with a decline in use was in the health sciences, which is not a core discipline in JSTOR, and another title with lower usage in 2011 had experienced publication delays, but, in general, CSP titles in core JSTOR subject areas were heavily used.

In response to a question, Muccie noted that JSTOR is actively working to expand the number of publishers in the Current Scholarship Program for 2013. JSTOR’s current approach is to target publishers with journals in core JSTOR disciplines with whom JSTOR already has a relationship through the archives and with titles that are not currently available online in a stable way. A number of librarians present mentioned that their libraries were actively seeking to cancel print subscriptions and to consolidate online subscriptions onto fewer platforms. One attendee encouraged JSTOR to continue to press publishers to offer small packages of online-only subscriptions at reasonable prices as this would be a welcome alternative to big deal packages from large publishers.

4. How Does One Measure ‘Value” with Long-Term Preservation? Looking at Portico After Five Years
The final presentation of the day was by Ken DiFiore (associate director, Outreach and Participation Services, Portico) who reviewed Portico’s evolution and future direction. Portico, founded in 2005, is a secure digital archive in which e-journals, e-books, and other electronic scholarly content are preserved. Its objective is to help libraries and publishers make a secure and reliable transition from print to online content.

Since libraries and publishers work together through Portico to preserve digital content, preservation costs for individual libraries and publishers are lowered. For libraries, Portico serves as an insurance policy for their investments in e-resources. Library members of Portico are guaranteed access to archived content, regardless of a library’s previous subscription status, when the content becomes lost, orphaned, or abandoned. These “trigger events” can occur when a publisher goes out of business, discontinues a title, or drops a back file. Some publishers also allow Portico to provide post-cancellation access; through Portico, libraries can access archival content for which they paid but to which they no longer currently subscribe. For publishers, Portico reduces or eliminates their internal archiving costs. Through Portico, they are able to meet library demand for perpetual access to content through a trusted, third-party archive without any impact on their own operations as Portico handles the conversion of source files to archival format and conducts future format migrations.

DiFiore explained Portico’s process for acquiring content. Portico enters into legal agreements with publishers, who commit to sending their content to Portico. Publishers pay Portico an annual fee, and they agree that content, once deposited, may not be removed. Portico is granted
the right to serve up content to its library supporters in the case of a trigger event, when the content is no longer generating revenue for the publisher.

When Portico was launched as a service in 2006, it included content from a dozen publishers. In 2011, Portico includes content from 133 publishers publishing on behalf of over 2,000 societies and associations. Of these publishers, 129 are journal publishers. Portico archives over 12,000 journals and can provide post-cancellation access for 89 percent of them. At this point, Portico has preserved over 16,000,000 articles, and the archive contains approximately fifteen terabytes of metadata and content. In recent years, growth of the archive has slowed as the ingestion of content from large publishers has been completed. Portico is now working to add content from smaller publishers with fewer titles.

Library participation in Portico is global with 705 participating libraries, about half of which are located outside the United States. For any interested library, Portico will conduct a free holdings comparison in which the library’s print and electronic journal holdings are compared with titles preserved in the archive to determine the degree of “insurance coverage” joining Portico would offer. DiFiore presented sample holdings comparisons for six institutions to illustrate that their preservation costs per title ranged between $1.08 and $2.17, which he stated is not much in comparison to subscription costs.
DiFiore next discussed recent changes to Portico’s service model. In January 2011, Portico began offering separate preservation services for e-journals and e-books as e-book content in Portico, and in libraries, was growing. Previously, libraries were required to support the preservation of all content types, and many libraries wanted more flexibility. Access scenarios for e-book access in Portico mirror the journal access model.

D-collections, on the other hand, work differently. Since typically only very large libraries have been able to afford d-collections (i.e., Gale’s Eighteenth Century Collections Online), Portico did not want to require small and medium libraries to support the preservation of these materials. Portico decided to adopt a different support model for these collections, whereby their preservation costs are fully supported by the publishers, not by libraries. Rules for access to d-collections are also different. Portico will provide trigger event access to previous library customers only, and post-cancellation access is not available. At this point in time, Portico preserves d-collections from Gale Cengage and Adam Matthew Digital, though they are engaged in active conversations with ProQuest, Readex, and Alexander Street Press about adding their content.

In the future, DiFiore explained, Portico is looking to preserve libraries’ locally-produced digital content, for example journals maintained on the Digital Commons platform. They are also hoping to partner with national libraries to satisfy national legal deposit laws.
At the conclusion of his presentation, DiFiore fielded a number of questions from attendees. In response to questions regarding the format and usability of digitized content, DiFiore explained that Portico normalizes ingested content as XML files, which are rendered to users as HTML. Portico supports OpenURL for accessing content and offers a basic search interface. Since their model is that of a dark archive, Portico will never offer the “bells and whistles” available on publisher sites.

Another question concerned e-book editions. An attendee from a medical library explained that books in the e-book collections to which her library subscribes are automatically updated to the latest editions when they become available from the publisher. Would earlier editions be accessible through Portico? DiFiore admitted that this was a good question and that currently an edition change does not constitute a trigger event. Older editions of e-books are maintained in Portico but are considered “inactive.” DiFiore echoed that Portico only serves up content when it is no longer generating revenue for the publisher and that this might be the case with older editions of e-books. He will look into this issue.

Another attendee asked why, as revealed in the downloadable list of e-journal holdings available from the Portico Web site, so many issues of preserved journals are missing (Portico, 2012). DiFiore responded that he was initially surprised to discover that publishers did not maintain their content in an organized fashion. He explained that Portico receives content from publishers that is “all over the place.” Current volumes of journals arrive with older volumes. Publishers have committed to Portico to send content, but not to which titles and volumes they will send and
when. He provided the example of Elsevier, which has over 9.4 million digital articles. It is a big effort for them to send a third party that much content.

Next, an audience member asked why a library would join Portico now as opposed to after it had lost access to purchased content. DiFiore replied, “What if every library took that position?” The audience member responded, “So altruism is the main reason libraries should support Portico?” DiFiore answered that there is a lot to be said for altruism as a reason to support Portico, but he believes that the progress Portico has made speaks for itself. Because libraries are investing so much in electronic resources, if they do not support Portico, they cannot be sure their investments are protected. DiFiore believes that there is a tangible business case to be made for Portico. He ended by stating that publishers are trying to do the right thing by committing their content to Portico. If they do not see libraries joining Portico, they will believe that libraries do not care about preservation, and the model will collapse.

References


Contact info:

Andrée J. Rathemacher
Professor
Head, Acquisitions
University Libraries, University of Rhode Island
15 Lippitt Road
Kingston, RI 02881-2011
Phone: (401) 874-5096
Fax: (401) 874-4588
E-mail: andree@uri.edu
http://www.uri.edu/library/