Smart Libraries through Technology

Disruption in Scholarly Publication

By Marshall Breeding

The realm of scholarly communications has entered a period of increased uncertainty, where pressures toward open access publishing increasingly present challenges to longstanding business models. The shift from the commercial subscription model to open access has major ramifications not only in the business economy for scholarly publishing but for the technologies involved in discovery and access to scholarly resources.

The digital age has been one of unfulfilled promises for scholarly communications. In the early days of e-journals, many believed that the transition from print would result in more immediate access to research findings and dramatically lower costs. It seemed at that time that rather mounting documents on web servers would involve almost negligible costs and could be accomplished rapidly. In some disciplines, this vision became more of a reality, as seen with the arXiv.org of physics preprints launched in 1991. But overall, the wholesale change from print publishing to electronic brought little change to the academic and business processes. The peer review process continued to be managed through commercial publishers and societies, and the business model continued to be based on subscriptions, funded primarily out of library budgets. The subscription model imposes significant barriers to access to academic research, given the broad disparities in the funding available across libraries to purchase subscriptions in each discipline. For items not covered by their library’s subscriptions, researchers can make requests through interlibrary loan or document delivery services, or pay the fees offered to non-subscribers. These “paywalls” represent a barrier for independent researchers not affiliated with a library.

The subscription model places a massive burden on academic libraries. The budgets of libraries, unfortunately, cannot support access to all possible resources available in all disciplines, making it necessary to be very selective about the journal titles to include in their body of subscriptions. The ongoing increases in subscription costs and the growth in new publications means that libraries must be very selective and make difficult choices regarding the resources they offer to their faculty and students. The need to carefully manage, track, and assess the investments libraries make in scholarly journals drives requirements for electronic resource management technologies, which were originally developed as discrete products and more recently incorporated into the functionality of library services platforms.

The pressure to mitigate costs and provide access to the widest possible range of scholarly content led to the emergence of the “Big Deal.” Many of the major publishers would negotiate a price with each university that would include coverage of all their e-journals. These deals, though costly, gave library users access to a very broad range of scholarly materials. These arrangements also meant that libraries were paying for many journal titles that they otherwise would not have selected. Libraries had to determine if the costs of individual titles they really needed purchased discretely would exceed the cost of the comprehensive offering, which might include considerable numbers of resources out of scope for the institution that may not see use. The Big Deal versus selective subscriptions has major implications for libraries.

In recent years, there has been some pushback with libraries not renewing their contracts and going back to discrete title selections. Regardless of the packaging, many academic libraries devote most of their collection budgets to subscriptions to electronic resources and have made deep cuts in monographs and other collection materials, as well as in other operational areas.

Open access presents an alternative to the subscription model. It has been an ideal model for scholarly publishing from
the beginning of the e-journal age but has proven difficult to realize. The costs of journal management, peer review, technical infrastructure, and other factors needed to support high-quality scholarly publications are substantial and require a sustainable business model. In contrast to subscriptions, where libraries pay for end-user access to journals, a commercial model of open access involves article processing charges (APCs) paid up front for published articles, which are then freely available for all to access. Some scholarly articles today are entirely open access; others are hybrid with a mix of restricted and open access. This business model shifts the financial burden away from the library to the researcher, and these fees are often covered by departmental funds or included in the budgets of grant proposals. There are other routes to open access, such as publishing agreements that allow preprints or accepted versions of articles to be made available on personal websites or institutional repositories.

After many years of slow progress toward open access for scholarly research, pressure has dramatically increased, primarily through the mandates of the organizations that fund research. In the United States, organizations such as the National Science Foundation and the National Institutes of Health require that articles produced from research that they fund be made available as open access. In Europe, several countries have issued mandates that all articles produced be made available as open access, and access to all content provided to their researchers for a negotiated national payment. In a bold move, eleven national research funding organizations in Europe, have asserted a new strategy, called Plan S that states "By 2020 scientific publications that result from research funded by public grants provided by participating national and European research councils and funding bodies, must be published in compliant Open Access Journals or on compliant Open Access Platforms."1

Such mandates and other pressures may move the scholarly publishing industry toward a tipping point where open access becomes the default model of publication. This change has broad implications for the resource management and discovery technologies used by libraries. The products available today, although they include capabilities to support open access materials, were designed primarily for the subscription-based economy.

On the resource management front, the business models of open access require different support than subscriptions. With open access, functionality would need to be built around article processing fees, whether the fees come from library or external budgets. In a hybrid environment, tracking access to scholarly resources moves form journal titles to individual articles, likewise for discovery services. The current model of index-based discovery is based on many assumptions related to the subscription model. These products now need to link to full text for all articles available as open access, even if the library does not hold a subscription for the journal. The task of managing and providing access to scholarly content would be simpler for libraries should comprehensive open access become a reality. That outcome seems unlikely, and distant at best, which means that the work of libraries in this will be much more complex than it was when almost all content was procured through subscriptions.

The changes afoot toward open access seem to be reshaping the business of scholarly publishing. Companies like Elsevier, Wiley, Clarivate, Digital Science, and SAGE, have each not only increased their involvement in open access, but have also shown a keen interest in the tools and technologies related to supporting institutionally conducted research and the workflow of scholarly publishing in addition to publishing the articles produced. These companies have made acquisitions or developed internally products related to peer review management, citation management, research data management, research information systems, discovery of research funding opportunities, scholars networks, and other activities. The library software vendors are likewise entering this arena, notably Ex Libris with its new Esploro product.

Open access naturally impacts library discovery services. The index-based discovery services are now well established among mid-sized and large academic libraries. These products continue to evolve and are increasingly able to lead searchers to open access content in addition to the resources covered within library subscriptions. While index-based discovery represents an important component of a library’s website, they reach only a limited portion of research activity. Many—if not most—library users rely on Google Scholar, Scopus, Web of Science, or other citation databases appropriate to specific disciplines. Tools have been developed to facilitate access to resources to assist users with gaining access to library resources or open access content when not using library-provided discovery services. A slate of browser plug-ins has been developed able to present the full text of an article from a website with its citation. These plug-ins work behind the scenes using resources such as the library’s link resolver or databases of open access articles to identify the source of the full text of an article whenever possible, often saving users from having to purchase it via the paywall.

We can anticipate additional tools and technologies to emerge as the realm of scholarly publishing evolves and open access content becomes more prevalent. The trend for

companies traditionally oriented to scholarly publishing to develop or acquire new businesses and their products oriented to scholarly workflows, discovery, analytics, and other related tasks will continue. Elsevier, Clarivate, and Digital Science have followed this pattern. This issue of *Smart Libraries* Newsletter features the expansion of yet another publisher, SAGE, as it expands into the realm of technologies in support of broader academic activities through the acquisition of Talis and Lean Library.

**SAGE Publishing Expands Its Tech Offerings**

There has been a strong trend in recent years for publishers to become involved with technologies supporting the broader scholarly communications process. SAGE Publishing has likewise joined this trend. Throughout its business history, SAGE Publishing has consistently expanded its portfolio of publications and its international involvement, both through launching new publications and through acquisitions. The company has increasingly been involved in open access publications. This year, SAGE has also ventured into the technology arena, through the acquisitions of Talis and Lean Library and with the launch of the Advance preprint archiving service.

**SAGE Acquires Talis and Its Reading List Management Application**

SAGE Publishing acquired Birmingham UK-based Talis, Ltd. in August 2018, bringing into the fold Talis Aspire, a reading list management application. Talis Aspire has been implemented in about 100 higher educational institutions, primarily in the UK, but with some installations in Europe and Australia.

Talis Aspire, launched in 2009, enables course instructors to incorporate scholarly resources in course packs and reading lists, primarily derived from the library’s body of subscriptions. These reading list management tools provide an important bridge between the library’s collection resources and the academic curriculum. In addition to providing a convenient way for instructors to assemble lists of the resources associated with a class, these products can reduce costs for students and the institution by making use of resources already covered by the subscriptions owned by the library before purchasing material from publishers.

Talis Aspire was the initial entry in this new category of reading list management applications. Other products in this genre include rebus:list, which was developed by PTFS Europe and sold to Kortext and the Copyright Licensing Agency in 2017, and Leganto from Ex Libris, which launched in 2015.

For SAGE, the acquisition of Talis means a closer insight into the way that scholarly materials are used in academic courses. Talis will join its Technologies for Learning and Research Division, led by Karen Phillips, Senior Vice President. For Talis, becoming part of SAGE Publishing provides opportunities to promote its product to a broader audience of academic institutions globally and to pursue more ambitious development. Under ownership by SAGE, Talis will continue to operate under its current leadership and its Birmingham, UK facilities.

**Talis Corporate History**

The acquisition of Talis by SAGE represents the latest phase of a very long and interesting history for the organization spanning almost five decades. The organization has operated under a variety of business models and has traversed through multiple computing eras and product offerings.

Talis traces its corporate history to the Birmingham Libraries Co-operative Mechanisation Project (BLCMP), established in 1969 as one of the early library automation initiatives in the United Kingdom. Early BLCMP members included the University of Birmingham and other major libraries in the surrounding region. It was organized as a non-profit cooperative, owned by its members and governed by a board of trustees.

The organization’s early services included cooperative cataloging service for its members, eventually leading to the development of automated systems operated on IBM mainframes. The capabilities of the BLCMP system culminated in about 1980 into a complete integrated library system. (Note: in the UK, the term library management system is used rather than integrated library system). An online catalog module was added by about 1985.

With the decline of mainframe-based computing, BLCMP began the development of a new library management system. This new Unix-based library management system named Talis was introduced in 1992 and offered as a commercial product for UK public and academic libraries.

In about 1999, the organization shifted from a non-profit cooperative managed by its trustees to a for-profit company named Talis after its flagship product. Dave Errington headed
the new company as its Chief Executive Officer. As a for-profit business, Talis was able to seek external funding, pursue business partnerships, and engage in other activities not easily accomplished when governed by trustees. At the time of this transition, its library management system was rebranded as Alto.

As a private, for-profit company, Talis diversified its development efforts to also include semantic web technologies in addition to the ongoing development and marketing of its Alto library management system and Prism web-based catalog interface. The company developed TalisList, the predecessor of Aspire. This product was initially launched in about 2000, following a development effort in partnership with the University of Sussex. The original TalisList product was not widely implemented.

During this period, Talis became a proponent of semantic web technologies and linked data for higher education and beyond. The company developed the Talis Platform as a strategic product, designed to manage business information in multiple business sectors using linked data. Talis developed a new version of its Prism discovery interface for its Alto learning management system as its initial application of its semantic web platform. Talis launched an Education Division in June 2009, with Aspire, originally introduced as TalisList, as its initial product.

The energy of the company was eventually shifted from library management systems to technologies surrounding linked data. In March 2011, Talis sold its library management division to Capita Group for around $32 million to fund its semantic web initiatives. Capita is a major provider of technology products and outsourcing services to local government authorities in the United Kingdom and saw the Alto library management system as a good fit within its portfolio of products for this sector. At the time of this acquisition, about 100 libraries had implemented Alto, all within the United Kingdom. The number of Alto implementations has decreased somewhat since that time, but it remains as one of the major library management systems in the UK. About 77 library services in the UK currently use Talis, which is also branded as Chorus when deployed as a hosted service.

Talis’ strategic focus on semantic web technologies proved to be a short-lived business activity. As noted in a Computers in Libraries article:

That bold move ultimately did not prove to be a huge commercial success. In July 2012, Talis announced its withdrawal from business activities related to the semantic web. The company, now somewhat downsized, focuses primarily on the Talis Aspire reading list management product for higher education and on digitization services. Despite growing interest in the semantic web, Talis acknowledged in its press announcement that commercial opportunities remain limited. It seems to me that although Talis was ultimately not able to build a profitable business model based on semantic web technologies, it was successful in building awareness of these concepts in the library community and beyond.

In the period since its exit from semantic web technologies, Talis saw considerable success in developing and marketing its Aspire reading list application. It has been adopted by most UK higher educational institutions as well by universities in Europe and Australia. The success of that product has culminated in its acquisition by SAGE Publishing.

**SAGE Acquires Lean Library**

In August 2018, SAGE announced that it had acquired Lean Library and its browser plug-in technologies that help researchers gain access to the full text of scholarly resources. Its products were designed with a library focus, helping to ensure a role of the library in resource access, even when the researcher does not start from library-provided catalogs or discovery services. Since only a limited proportion of students and researchers affiliated with a college or university begin their research using the discovery services provided by the library, it is important to provide alternatives for authenticating and linking to subscribed resources outside those environments.

The May 2018 issue of Smart Libraries Newsletter included a brief summary of Lean Library and its products:

Lean Library offers a browser extension branded as Library Access that facilitates access to scholarly resources. It emerged from a prototype originally developed at Utrecht University Library, which gained distinction through its strategy not to offer its own discovery service but to encourage the use of Google Scholar and other disciplinary resources.

The Library Access browser extension from Lean Library facilitates access to scholarly content provided by the library. The extension sits in the background until the user opens a resource available with the library’s subscriptions, and then it presents the versions of the link from the institution’s proxy server. Library Access works to make it easier to access library-provided resources, removing any complications related to authentication from on-campus and off-campus locations. Like Unpaywall and Kopernio, it attempts to find open access copies when subscribed versions are not available.

Lean Library focuses on the perspective of the library to assist its users in gaining easier access to subscribed resources and to provide analytics related to usage that can be used to inform decisions on developing its collection of electronic resources. Lean Library appeals to libraries through its policy of not sharing or selling usage data. The data it collects is anonymized and not shared beyond the library.

Lean Library was founded by Johan Tilstra in October 2016 with assistance from UtrechtInc, an incubator that provides services and financing to new startups. Prior to launching Lean Library, Tilstra was affiliated with Utrecht University.3

In addition to Library Access, Lean Library has also developed Library Alternatives, which provides supplementary functionality to identify other copies of an article of interest that might be available in open access repositories or from within the library's portfolio of subscriptions. Library Alternatives works with the organization's link resolver behind the scenes.

Library Assist provides site-specific messages from the library displayed on the user's web browser. The library, for example, can configure Library Assist to display a message when the user connects to the website of a publisher for which a subscription is available. The message might notify the user that the library has purchased a subscription and provide information on obtaining any needed login credentials.

Browser plug-ins have been a product genre gaining increasing interest in the scholarly communications ecosystem. The May 2018 issue of Smart Libraries Newsletter included coverage of the acquisition of Kopernio. Unpaywall, developed by the non-profit Impactstory, falls within the same niche as Kopernio and Lean Library.

### SAGE Launches New Advance Preprints Service

SAGE has launched a new open access preprint service for papers in the social sciences and humanities, Advance. Positioned as a preprints community, Advance serves as a repository for research papers that have not yet been peer reviewed or published. Advance facilitates early availability of research as well as a vehicle for submission of papers to appropriate SAGE journals. Preprint servers such as Advance not only provide early access to research but also ensure that authors receive credit for their work. All papers submitted to Advance are assigned a DOI and remain permanently in the repository even when they are subsequently published.

The Advance service has been implemented in partnership with Figshare, which provides its technology platform. Figshare's primary service is a repository launched in 2011 for datasets, images, video, and other files produced through research projects. Figshare is owned by Digital Science.

SAGE offers Advance as a free service. Preprints deposited in the repository can simultaneously be submitted to SAGE journals, but there is no obligation. The platform provides a commenting feature so that authors can receive feedback from readers. Preprints in the repository are available as open access and can be cited and shared freely. This service aims to benefit the academic community in the social science and humanities in the same way that arXiv supports disciplines in physics. SAGE benefits from Advance as a streamlined submission channel for its journals.

### SAGE Publishing Background

SAGE has grown from a small firm established in 1965 issuing a single journal to one of the largest academic content businesses with a broad and diverse portfolio of publications and multiple international divisions. The company was founded by Sara McCune who led the company as its President through 1989 and as its Executive Chairman thereafter. In 2004, Blaise Simqu was named Chief Executive Officer of SAGE.

SAGE is privately owned by Sara McCune, with a succession plan to transition to ownership by a charitable trust to ensure the perpetual independence of the company.


The company has also made advancements in technology consistent with evolving trends. The company was launched in the era of print publishing. As the industry shifted to electronic publication, so did SAGE. Beginning in about 2000, SAGE began the transition of its journals to an online

---


SAGE has also seen considerable geographic expansion. Beginning from its original formation in New York in 1965, the company relocated to California in 1966 and has moved its headquarters facilities multiple times to accommodate its growing workforce. New international divisions were formed, including SAGE Publications, Ltd (London, 1971), SAGE Publications India (1981), and SAGE South Pacific (2006). Today SAGE employs over 1,500 personnel and is active on five continents.

SAGE also created business units and partnerships as it became increasingly involved with technology. These endeavors included Scalari software division (1995-2004) and a partnership with Acrobatiq for adaptive courseware in social sciences (2014). The acquisitions of Talis and Lean Library further expand the company’s technology portfolio.

### Smart Libraries Q&A

Each issue, Marshall Breeding responds to questions submitted by readers. Have a question that you want answered? Email it to Samantha Imburgia, Associate Editor for ALA TechSource, at simburgia@ala.org.

**What kind of disaster response plan should we have in place in regards to data management and security?**

Protecting data remains a top priority for libraries, especially since they rely on technology for almost all aspects of their operations and service delivery. Interruptions due to technology failures are not only inconvenient to library users and personnel but also harm the library’s reputation. All technologists should be well versed in the core precepts in disaster planning and recovery and institute proactive measures for their organization. It’s essential for libraries to safeguard their data against any possible failure in hardware, software, or human error.

The relative proportion of risks have changed dramatically over the course of my career in library technology systems. In earlier times, most failures were related to hardware failures. Hard drives, storage arrays, and other devices were prone to failures, making it important to ensure that all data were copied as frequently as possible to another medium. Today, organizations are less dependent on local hardware components, with greater involvement with highly-redundant cloud-based storage services. Malicious attacks and human error represent some of the worst threats to data security.

Any effective disaster plan must work toward minimizing any operational impact potentially incurred by any type of technology failure. These strategies are based on a design of technical infrastructure able to remain operational even if one or more components fail and to prevent loss of data even in the event of failure. To achieve these goals, systems need to have multiple layers of redundancy, data synchronized across multiple systems or repositories. Ideally, organizations will monitor systems for potential issues before they result in data loss and have automated and human processes to recover data and systems when failures happen.

Data protection depends on redundancy and replication across devices so that the operational systems can withstand failures of underlying components. In the earlier days of storage, organizations often depended on RAID (redundant arrays of independent discs) technology to be able to remain operational, sometimes with degraded performance, even if a disk drive component fails. Despite this initial layer of protection, storage devices could fail, often related to errors in controllers, drivers, or software causing wholesale data corruption. RAID continues to be a viable storage technology, though many other alternatives are implemented in large-scale data centers.

Given the inherent vulnerability of primary storage devices, organizations necessarily implement back-up schemes involving the regular transfer of data to secondary, usually offline, media. Secondary media are often kept offsite as an additional safeguard in case of catastrophic events at the organization’s primary business location. Systems administrators often implement software designed to automate backup procedures, including the performance of incremental and comprehensive data backups, rotation of media, and routing of offsite copies.

The era of cloud computing has drastically altered the nature of disaster planning and recovery. Many organizations have shifted from local computing and storage equipment to cloud services or to hybrid environments. Cloud technologies provide many options for security that can provide even
greater protection for data than were possible through traditional processes used for on-premises infrastructure. Many organizations now rely more on outsourced or contractual enforcement of data protection strategies than hands-on backup procedures.

Organizations increasingly rely on cloud-based data management and storage services to house their operational data. Amazon’s S3 (simple storage service), for example, provides a highly reliable and widely used service for data storage. These cloud-based storage services implement multiple layers of hardware and software redundancy that automatically work around component failures, usually with no interruption of service or performance degradation.

Although cloud-based storage services have incredibly low failure rates, they must also be supplemented by additional layers of data protection. In the same way that an organization would never keep only one copy of its data on a single local storage device, multiple layers of redundancy should also be implemented when relying on cloud-based storage services. Many organizations will deploy data strategies with simultaneous replication to multiple storage services in separate data centers, ideally in multiple geographic regions. Such a data architecture can be designed to withstand widespread component failures within a data center or even the complete loss or unavailability of an entire facility.

Other precautionary measures include routine archiving of data to offline storage services, such as Amazon Glacier. These measures support both disaster planning and recovery as well as organizational requirements for archiving and data retention. Archived copies of database files, for example, may be needed to satisfy requests for records active in a previous period, but since deleted from the active business applications.

Disaster planning in a cloud environment can also take into consideration business issues related to the service providers. Some organizations may want to make replicates of data on storage services from multiple providers. This layer of protection may incur significant cost, including duplicate storage service subscriptions as well as for connectivity fees. It is also possible to implement automated processes that regularly transfer data to on-premises storage devices. These measures ensure ongoing access to data in the event of a business failure or an account or contract dispute.

Both local and cloud-based data strategies can be crafted and implemented to protect operational data against almost any technical failure. It is much more difficult to guard against human error and malicious attacks.

Human error can introduce corruption or loss of data that can be extremely difficult to prevent or remediate. In the context of library systems, for example, a script or command to perform global changes can introduce widespread errors. Such problems can be especially difficult to repair if they are not detected quickly. Beyond a given threshold, errors will be propagated throughout the online and offline replicates of the databases involved. Protection against these scenarios requires careful testing of all procedures that update operational databases and produce replicate or backup copies.

The phenomenon of ransomware can also present quite a challenge to data protection strategies. Introduced through malware, these attacks attempt to encrypt the data of the organization, often including critical operational databases. Once encrypted, only the attacker would hold the digital key needed for decryption, which would be provided only if the ransom demand is met, usually in the form of large payments made via Bitcoin or other non-traceable currency. Should the encrypted version of the database be propagated to replicate and backup copies, recovery apart from paying the ransom can be complex and sometimes impossible. Protection against this type of attack can be accomplished by making copies of critical data on devices not directly connected to the filesystems or business applications of the primary business environment and implementing procedures that test for unauthorized mass encryption.

As libraries increasingly depend on vendor-hosting for their core business applications, such as their integrated library system, library services platform, or digital collections management applications, that involve critical operational data, it is important to understand the redundancies in technical infrastructure and the disaster recovery procedures instituted. Any contract or subscription agreement for a hosted service will include a service level agreement (SLA) specifying these details. The SLA will state the required level of system availability and penalties for excessive downtime. It should also detail the procedures to protect customer data, including operational replicates, offline backups, and processes to provide copies of the data upon termination of the service contract or other triggering events.

Disaster planning and recovery strategies remain an essential component of a library’s technology strategy. The latest round of changes toward cloud technologies have brought significant changes, with a new set of pragmatic methods to accomplish the basic principles of data redundancy and easily recovered backups. Data management strategies must be reviewed periodically to ensure they remain viable as technologies evolve and as the organization makes changes in its infrastructure.
October 2018
Smarter Libraries through Technology

Smart Libraries Newsletter

Marshall Breeding’s expert coverage of the library automation industry.

Editor
Marshall Breeding
marshall.breeding@librarytechnology.org
Twitter: @mbreeding

Managing Editor
Samantha Imburgia
312-280-3244
simburgia@ala.org

Digital Access for Subscribers
journals.ala.org/sln

TO SUBSCRIBE

To reserve your subscription, contact the Customer Service Center at 800-545-2433, press 5 for assistance, or visit alatechsource.org.

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.