

White Paper

Digital Preservation:
The Achievements, Challenges and Encouraging
Future

Read
Now



Ultimately, librarians must be able to rely on the long-term availability of content in digital form for many decades in the future

Digital Preservation

The Achievements, Challenges and Encouraging Future

The digital preservation of scholarly resources has become critical over the last 15 years since the first electronic journals appeared in the 1990s. As electronic access has become the format of choice for most institutions, librarians have recognized the potential perils of investing in electronic content and leaving behind traditional print materials. Without an established plan for long-term electronic access from the outset of this transition, both librarians and publishers faced questions that are now being addressed by regional and worldwide digital preservation initiatives. Some of these concerns include problems such as:

- What happens to the existing content if a journal is no longer published?
- What can one do with online journals when a publisher goes out of a business? Who is responsible for hosting the published content?
- What happens if the hosting server is no longer functional?
- What happens if our current computers go the way of the floppy disk?

These are not merely questions addressed with a perpetual access clause for individual libraries at the point of purchase. Ultimately, librarians must be able to rely on the long-term availability of content in digital form for many decades into the future.

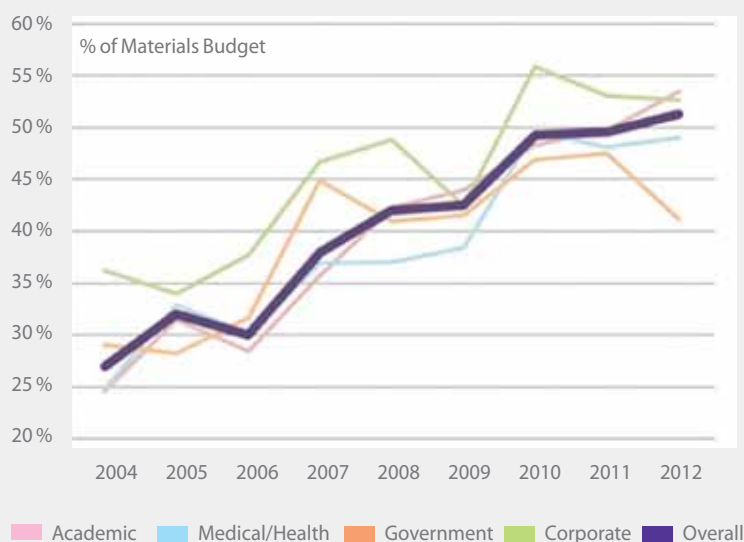
Springer and Publishers Communication Group have explored the global views of digital preservation of scholarly content. In addition to independent research, industry experts were interviewed, including representatives from three of the largest digital preservation initiatives (Portico, LOCKSS and CLOCKSS) to provide insight on what has been achieved and what we may expect to see in the future. Several librarians from across the globe have contributed feedback about their libraries' activities with digital preservation and why they value participation in preservation efforts.

The intention of this study is to highlight the importance of digital preservation of scholarly content, the achievements that have been made in this endeavor and the on-going efforts from both third-party organizations and publishers such as Springer to seek an effective solution.

The Importance of Digital Preservation

Prior to the explosion of interest in online scholarly resources, the print format for scholarly output allowed libraries to archive their own resources relatively easily. However, the advantages of electronic access for researchers and libraries alike have led to very rapid adoption of digital content despite concerns about digital preservation. As seen in Figure 1, in the last nine years the proportion of electronic materials to physical materials has continued to trend upward in nearly all segments.

Figure 1: The estimated amount of the materials budget spent on electronic information sources for libraries around the world has been trending upward over the last nine years



Source: 2013 Annual Library Budget Survey, Publishers Communication Group



In the last nine years the proportion of electronic materials to physical materials has continued to trend upward nearly in all segments

Early in this shift toward digital, libraries continued to acquire print resources whenever possible in order to maintain an archive for their institutions. Soon, however, online subscriptions were including supplemental materials that were simply not available in the print versions. Publishers eventually created pricing structures significantly favoring the electronic subscriptions or ceased production of the print format entirely. Libraries could no longer ensure a complete archive of the published content in their print collections and were now forced to rely on the comparably fragile online archives and perpetual access agreements with their publishers. Roxanne Missingham, Chief of Scholarly Information Services at the Australian National University, articulates the importance of finding a solution: “the research produced by universities is of international significance and needs to be made available permanently through well managed digital archives... The resources for research and teaching are vulnerable.”

Libraries invest thousands of dollars in their subscriptions and purchases each year but risk losing access to these investments in the future unless robust long-term solutions are available. Despite the best intentions of publishers, unforeseen circumstances may lead to the loss of access at some point in the future. This means that content which has already been purchased and which is relied upon by researchers can be lost due to outside factors including economic downturns, natural disasters, and widespread technical blackouts. Librarians and publishers quickly began to realize the need to design an infrastructure capable of long-term digital preservation. Ultimately, the recognition of the fragility of online content led to the digital preservation movement and the creation of several initiatives tasked with the responsibility of overcoming these obstacles.

Major Digital Preservation Initiatives

This overview explores the development and impact of three major international digital preservation initiatives: Portico, CLOCKSS and LOCKSS. These three organizations have had significant impact in the preservation effort, each with its own approach. A comparison of these initiatives will show that there are some differences in presentation, funding, and structure – but all complement one another in the effort to create reliable, more complete long-term archives. To facilitate this comparison, representatives from each of these three initiatives were interviewed, including:

- Victoria Reich, Executive Director and co-founder of the LOCKSS program
- Randy Kiefer, Executive Director of the CLOCKSS Archive
- Kate Wittenberg, Managing Director of Portico



The LOCKSS Program

Lockss.org

The LOCKSS Program, or “Lots of Copies Keep Stuff Safe”, is an open-source, library-led “light” archive digital preservation system. As the name suggests, LOCKSS relies on a distributed network of libraries to preserve multiple copies of published content in their local “LOCKSS box”. This technology ensures that users will have immediate and continuous access to publishers’ content whenever it is needed. LOCKSS offers a preservation program that builds library infrastructure and collections, which in essence mirrors the traditional purchase-and-own library model.

Victoria Reich and David Rosenthal founded LOCKSS with a grant from the National Science Foundation in 1998 and soon after received major support from Sun Microsystems, the Mellon Foundation and the Library of Congress.



CLOCKSS

Clockss.org

The CLOCKSS Program (Controlled LOCKSS) was built off of the efforts initiated by LOCKSS in equipping libraries with LOCKSS box technology. However, the function of CLOCKSS is different in that it is a “dark” archive. While LOCKSS provides libraries with the ability to ensure uninterrupted access to a publisher’s content, CLOCKSS content must be “triggered”. Material held in CLOCKSS may be made available only when there is a loss of access from all sources, and the Board votes to trigger the release of the content into open access.

CLOCKSS was founded in 2006 and was incorporated as a charity in 2009. Randy Kiefer became the Executive Director in 2011.



Portico

Portico.org

Portico is a digital preservation service provided by ITHAKA, a not-for-profit organization that is also responsible for the short-term access scholarly journal archive JSTOR. Like CLOCKSS, Portico is a “dark” archive that aims to preserve the usability, authenticity, discoverability and accessibility of academic content for long-term preservation of scholarly content. Portico relies on proprietary software that archives publisher-provided metadata, rather than the LOCKSS box technology on which both CLOCKSS and LOCKSS rely. Content held in Portico is only released after particular “trigger” events, at which point it is made available to Portico participating libraries.

Created in 2002, Portico began as a project funded by the Mellon Foundation. Kate Wittenberg became the Managing Director in January 2012.



Type of Content

The type of content being preserved by these initiatives has few limitations. Portico preserves journals, books and primary source content, including open access publications. Kate Wittenberg explained that Portico currently does not preserve blogs or other “informal forms of scholarly communication”, though they are not limited to specific file formats. They also include a wide range of audio, visual and more esoteric discipline-specific files. This is also true of CLOCKSS, as Randy Kiefer states that the content preserved by CLOCKSS is “essentially any form of web publishing associated with formal academic content”. As for LOCKSS, the structure allows for the determination of preserved content to be made on the librarian level given the permission of the publisher and existing access for the institution, though the content preserved currently is similar to that of the other organizations.

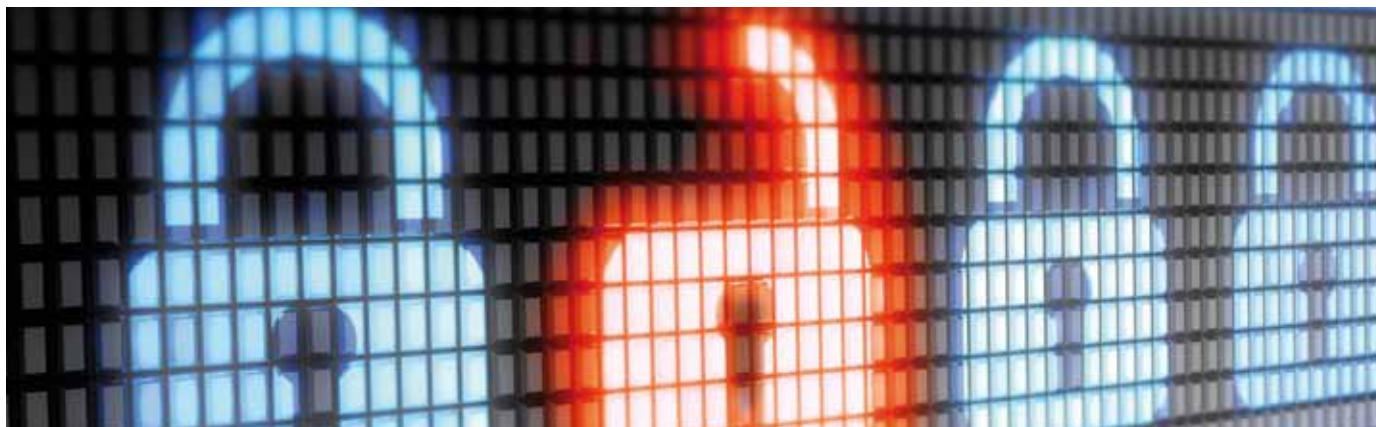
Types of Archives

There are two types of archives that are typically discussed in long-term preservation, known as light archives and dark archives.

A light archive, such as the archiving done by LOCKSS, is content that becomes available to users when a publisher’s website is temporarily unable to provide access. The archive is once again inaccessible when the publisher recovers the ability to supply their content to the reader. The intention of this type of system is to ensure that the user will receive 100% continuous access.

A dark archive, such as Portico and CLOCKSS, is inaccessible (or “dark”) until a particular event occurs that makes the content unavailable in any other way. Once one of these “trigger” events occurs and is identified as such by the governing body of the archive, the content is made available and the archive is presumed to be the sole digital provider of this information.

Both types of archives differ from short-term archives, which are associated with post-cancellation rights. The archives described above are intended for long-term, last resort access to scholarly content.



A dark archive, such as Portico and CLOCKSS, is inaccessible (or "dark") until a particular event occurs that makes the content unavailable in any other way

Trigger Events

CLOCKSS and Portico, the two dark archives included in this study, only make content accessible once a specific trigger event has occurred. A "trigger" event is the loss of access from any source to digital content. This may mean that:

- Publisher goes out of business
- The hosting or publisher platform experiences catastrophic failure for the long-term
- The journal ceases publication and digital access is removed

In some situations, a journal may no longer be available through the publisher's site, but still remains accessible via another site such as an aggregated database. This would not be considered a trigger. Dark archive access is the source of last resort, and triggered content must go through an approval process before becoming available. Portico works with publishers to identify when these trigger events occur and alerts participating libraries to the new access method as quickly as possible. CLOCKSS uses a 24-member board consisting of librarians and publishers that votes on whether to release the content. The board must have the approval of at least two-thirds of the members (16 votes) with no more than two negative votes. Currently, this process takes about 10 days from the loss of access for CLOCKSS to release content from the archive.

Accessibility

Once content is made available through any of these archives, it is important to note the differences in who will be able to access the resources. In the case of LOCKSS, member libraries have transparent access to the content preserved in their own institution's LOCKSS box. Libraries change what subscribed content (or openly-accessible content) they wish to archive. When content is made available, the library may access their locally stored copy. When a "dark" archive title is triggered through CLOCKSS, the content becomes open and free for all, regardless of library participation. Portico differs from this approach and provides access and support to only Portico-participating libraries.

Funding and Structure

The accessibility of content is in part related to the structure of these initiatives. LOCKSS is funded by libraries participating in the LOCKSS Alliance. LOCKSS provides free open source software to member libraries. Using this software, libraries create a "LOCKSS box" that stores their own local archive of content. Publishers may participate in LOCKSS for no fee. A Mellon Grant for strategic development also supports the organization.

CLOCKSS charges a participation fee for publishers. Libraries may voluntarily contribute funds to support the program, but otherwise are not charged.

Portico is funded by fees from participating libraries and publishers. It has the infrastructure to provide user support for triggered content, as well as additional preservation and curation services as part of its archival management.

Content Acquisition and Presentation

Content acquisition is another major differentiating factor for these organizations. Portico works directly with publishers' production-level content and repackages their source files into an archival format. This includes an initial transformation of the publisher-provided descriptive (bibliographic) metadata upon its intake into the archive. The difference of this system is apparent in the presentation of content to the end user. Content preserved by Portico has a uniform look and feel, and all user support is provided by Portico

LOCKSS uses its open source software to crawl publisher content directly from their website. The publisher's layout, branding, look and feel is preserved and presented to the end user to facilitate a seamless transition should content be access be needed. Publisher websites continue to receive hits via the LOCKSS box when it is providing the access. Vicky Reich explains, "Publishers spend considerable resources thinking about how best to present their content to readers. It deserves to be preserved."

Likewise, CLOCKSS utilizes LOCKSS software to preserve content. Therefore, the content retains the original appearance of the published material from the publishers' site. Content for CLOCKSS is maintained by twelve libraries around the world, referred to as "archive nodes", rather than a few proprietary servers.

Adjusting to Change

Plans for long-term preservation must also account for a variety of future changes in technology. For CLOCKSS, Randy Kiefer describes it as an "ever-changing challenge." Both LOCKSS and CLOCKSS benefit from the open source LOCKSS software, as engineers around the world continue to develop and improve performance at a fairly low cost. Additional grant funding also allows for continued development to adjust for migration of content to new formats as they develop.

Portico takes a migration-based approach to preservation to ensure long-term usability of content. Portico has also developed automated tools to verify the accuracy of these formats and regularly participates in research efforts focused on the requirements for long-term accessibility of different file formats, including open source tools.

Ease of Participation and Use

Librarians looking to participate in digital preservation initiatives, whether working with one or more of these organizations, should be aware of what is required to begin and what will be needed to maintain participation into the foreseeable future. With Portico, there is a trade-off between cost and ease of participation. While the library is required to pay a fee to participate, customers receive on-going customer support along with the ability to access all triggered content. Beyond the annual participation fee, little is required of the library.

LOCKSS also requires an annual fee for libraries. The institution then receives LOCKSS Box software and technical support to bring the software online and create their own localized archive for their materials. Mark Jordan, head of library systems at the W.A.C. Bennett Library of



"Publishers spend considerable resources thinking about how best to present their content to the readers. It deserves to be preserved."

Victoria Reich, Executive Director and co-founder of the LOCKSS program



Simon Fraser University in Canada, states that, “the physical and technical resources required for participation in LOCKSS networks are nominal and are a small price to pay for libraries who want to maintain direct control over their preserved digital assets.”

CLOCKSS is funded through voluntary contributions from libraries and a participation fee from publishers. Libraries must certify that they have security provisions in place to protect publisher content. However, there is no further action required to take part unless the library is one of the twelve institutions maintaining a CLOCKSS archive node via their LOCKSS software.

Summary

This brief comparison of these multi-faceted organizations shows the breadth of work being done to facilitate long-term digital preservation. Figure 2 summarizes these points in a simple reference chart, while further information can be found at each organization's respective website.

Figure 2: Comparison of key differentiators of digital preservation initiatives Portico, CLOCKSS and LOCKSS

	Portico	CLOCKSS	LOCKSS
Type of archive	Dark	Dark	Light
Accessibility	Last-resort access when no other source will be made available	Last-resort access when no other source will be made available	Provides temporary, seamless access to facilitate 100% perpetual access when a publisher is unable to provide access
Who can see it?	Portico member libraries may access triggered content	Anyone may access triggered content	Participating libraries can access content available on their local LOCKSS box
Fees	Fees for library and publisher	Fee for publisher and voluntary contribution from library	Fee for library
Presentation	Uniform look across all preserved content	Mirrors publisher presentation	Mirrors publisher presentation
What libraries need to do	Become a member	Group of libraries act as archive nodes, otherwise no action required	Acquire and implement LOCKSS Box

Regional Initiatives

In addition to the three initiatives outlined above, local and regional efforts that have proven to be successful and influential are in place around the world. Examples of these can be found in libraries in North America, Europe, Asia or Australia that are actively implementing solutions to address the question of long-term preservation. Regional initiatives tend to focus on content published locally, typically aiming to preserve content according to the country of origin. These projects complement the broad international efforts taken on by organizations such as LOCKSS, CLOCKSS and Portico.

The importance of all of these initiatives is exemplified by many libraries participating in preservation efforts at both international and local levels. For example, the Università Cattolica del Sacro Cuore (UCSC) in Milan acts as a CLOCKSS archive “node” and also participates in national preservation initiatives. Ellis Sada, Head of the Library, explains her institution’s perspective: “We consider initiatives at the international level essential. However, they must be coupled with a national strategy that we are actually fostering. Italy is an important player in scientific and scholarly research and in cultural studies; therefore it has to play a leading role in digital preservation.”

In the Netherlands, the Koninklijke Bibliotheek (KB), or National Library, was one of the first to create a digital library and set up preservation agreements in the 1990s. Acting as a center of expertise for digitization and digital preservation is one of its major activities. The e-Depot is the long-term repository created to help accomplish these preservation goals. According to the library’s website, “in and of itself the KB’s e-Depot is neither a dark archive nor a light archive.” Access to content depends on each archiving agreement. Users willing to pay for an annual library pass can access all content from the archive for which the KB has user rights, though generally only for on-site use. The library plans to make access simpler for a greater number of users through partnerships with the academic community and university libraries.

The Deutsche National Bibliothek, or German National Library, is another early adopter of long-term digital preservation policies. In its role as the national archive library, it partnered with the Niedersächsische Staats- und Universitätsbibliothek Göttingen and other institutions in a project known as kopal in the mid-2000s to set up a digital repository with the aim of long-term preservation. The goal, according to the kopal information site, is not only preservation of the DNB’s resources, but also “reuse by other cultural heritage institutions and by other institutions that need long-term archiving is expressly desired.”

Along with these two examples, many other regional initiatives have undertaken substantial research to improve their own systems and share their knowledge with libraries.

These are just a few examples of the wide range of regional initiatives in progress to preserve national and international work for the long-term. The majority of these libraries are working in partnership with universities and other members of the academic community within their region to facilitate long-term preservation efforts and continue to improve their technical infrastructure.



"We consider initiatives at the international level essential. However, they must be coupled with a national strategy that we are actually fostering."

Ellis Sada, Head of Library, Università Cattolica del Sacro Cuore



Many regional initiatives have undertaken substantial research to improve their own systems and share their knowledge with libraries

Figure 3: A sample of regional initiatives representing the significant efforts happening around the world

Institution	Program	What is it?
Swedish National Library	National Archive	The National Library is developing a technical infrastructure for a digital archive and to ensure long-term availability. The archive will house digitized material and collected digital material. The archive consists of two parts, a "dark archive" that provides a secure storage site for original files, and the open archive (digital repository), which is adapted for the users.
National Library of China	Digital Library Project	The goals of the DLP are twofold: first, to preserve in digital form manuscripts and books embodying China's history and cultural heritage; and second, to capture and preserve for the long-term more ephemeral forms of today's dynamic Chinese culture – Web pages, e-journals, blogs and so on – that appear only electronically.
National Library of Australia	PANDORA	This archive was established initially in 1996 and was built in collaboration with nine other Australian libraries and cultural collecting organizations. This effort led to the development of the PANDAS system (the technical system supporting the PANDORA archive), selection criteria and other infrastructure components to support the capture and preservation of Australian publications online.
Library of Congress	NDIIPP	The National Digital Information Infrastructure and Preservation Program (NDIIPP) is implementing a national strategy to collect, preserve and make available significant digital content. The program has engaged hundreds of organizations and partners across the United States and around the world. A major current initiative is the National Digital Stewardship Alliance.
National Library of Japan	National Digital Library	Since 2002, the National Digital Library has been conducting studies for ensuring long-term preservation and usability of digital information.

The Role of Springer in Digital Preservation

Springer participates in a number of digital preservation initiatives, including: Portico, LOCKSS, CLOCKSS, the German National Library and the Koninklijke Bibliotheek. As a large global publisher, Springer recognized the importance of participating in these programs to help influence the involvement of smaller publishers and to demonstrate the significance of long-term digital preservation. In fact, Springer was one of the pilot participants in the CLOCKSS initiative. Heather Staines, formerly the Senior Manager eOperations for Springer, explains, "Springer looks to preserve everything they can. They want to be responsive to customer demands and make participation become an industry standard." Figure 4 demonstrates the extent of Springer's involvement with preservation, with over 1500 titles approved for CLOCKSS and Portico. BioMed Central open access titles are also included in these archives, though not reflected in the count.

Figure 4: Springer's participation in digital preservation initiatives as of July 2012

Digital Preservation Initiative	Springer-owned Journals	Third Party, Society, Distribution Journals	eBooks	Archives	Reference Works	Post-Cancellation Access
Portico	983	631	Subject Collections + book series	For approved titles	Yes	Yes
	983	631	Subject Collections + book series	For approved titles	Yes	n/a
LOCKSS	Matches license	Matches license	Matches license	Yes	Matches license	Yes
GNL	All titles/all volumes since 2007	All titles/all volumes since 2007	Titles publishing since 2007	No	Yes	At the GNL only
KB	All titles/all volumes since 2005	All titles/all volumes since 2005	Titles publishing since 2009	No	Yes	At the KB only

Source: Digital Preservation: What You Need to Know, presentation by Heather Ruland Staines, eOperations for Springer, Heidelberg, July 23, 2012



Ongoing Importance of Digital Preservation

As shown by the examples highlighted in this paper and the many other initiatives underway around the world, long-term preservation is a critical issue. More and more libraries are actively seeking solutions, and publishers, in turn, are taking steps to ensure preservation of their content.

For publishers, having the security of a digital preservation plan goes well beyond a marketing statement and has become a key responsibility born by disseminating key research for generations to come. By the same token, librarians' participation is important to demonstrate to both publishers and authors that the responsibility of long-term preservation is indeed a collective effort. Echoing what is discussed in many conversations, Portico's Kate Wittenberg agrees, "Publishers and librarians do need to share the responsibility and costs of digital preservation as it is critical for their missions and their users."

The collaborative approach between libraries and publishers such as Springer has been a positive element in this endeavor and will continue to be an essential component in developing robust solutions. However, all sides agree that continued innovation is crucial to increase efficiency and capacity for the ever-growing body of published digital content. Fortunately, regional projects coupled with international initiatives reveal that a great deal has already been achieved in the digital preservation effort. These pioneering accomplishments have laid the groundwork for a promising future and are ensuring that the scholarly publishing community stays both active and engaged in this process.



Author

Kate Lara

Market Research Manager at Publishers Communication Group

Special thanks to contributors:

- **Randy Kiefer**
Executive Director of the CLOCKSS Archive
- **Victoria Reich**
Executive Director and co-founder of the LOCKSS program
- **Kate Wittenberg**
Managing Director of Portico
- **Heather Ruland Staines**
Vice President, Publisher Development of SIPX
(previously Senior Manager eOperations for Springer Science + Business Media)
- **Ellis Sada**
Head of the library at the Università Cattolica del Sacro Cuore
- **Roxanne Missingham**
University Librarian (Chief Scholarly Information Services) of the Australian National University
- **Mark Jordan**
Head of library systems at the W.A.C. Bennett Library of Simon Fraser University
- **Michael Seadle**
Professor and Director of the Institut für Bibliotheks- und Informationswissenschaft at Humboldt-Universität zu Berlin

Would you like to know more? Contact Springer today.

Visit springer.com/salescontacts to find your local Springer representative!

Springer

» New User
LOGIN

HOME | MY SPRINGER | SUBJECTS | SERVICES | IMPRINTS & PUBLISHERS | ABOUT US

» *Sales Contacts* Home » For Librarians » Sales Contacts

Search... GO
Advanced Search

Contact Us for Library Sales and Marketing Questions

Academic Librarians

- » United States
- » Canada
- » Central and South America, the Caribbean
- » Europe & Africa
- » Asia, Middle East & Australia

Corporate & Health Librarians

- » The Americas
- » Europe & The Rest of the World

Government Librarians

HOW TO CONTACT US

- Send us an email
- Online Service Helpdesk
- Customer Support
- Library Relations

Contact us!