



Smart Libraries™

Formerly Library Systems Newsletter™

50 East Huron Street, Chicago, Illinois 60611-2795, USA



Smarter Libraries Through Technology

Beyond IP Authentication: The Need to Modernize Access to Library Resources

By Marshall Breeding

Providing access to electronic resources is a core role of libraries. Academic libraries devote most of their collection budgets to subscription fees for these resources and therefore have a critical interest in making sure that they can be easily accessed by their patrons. Many of these resources cannot be accessed freely by the general public on the web but must be restricted to those associated with a subscribing institution. Those not associated with a library subscribing to the resource might be able to purchase a personal subscription or pay for individual articles. Although an increasing portion of the new scholarship is being published as open access, the model of restricted subscription-based resources remains a fundamental part of the library information ecosystem and requires a technical solution.

IP Authentication

The longstanding technical approach for enabling access to restricted resources for individuals associated with subscribing

has been based on IP authentication. Basically, if a person connects via a device associated with one of the network addresses assigned to an institution that has purchased a subscription, then access is allowed. Access from any other address encounters a paywall, a page that offers other alternatives to view the article, such as paying a per-article charge or prompting for an individual or institutional login.

IP authentication assumes that a person's institutional identity can be associated with the IP address of the device used to access the content. Each device must have a globally unique IP address and blocks of these addresses are assigned for use by the networks of educational institutions. The scheme of using IP addresses to validate access to restricted scholarly content worked well in the early phase of electronic resources but has been problematic ever since.

This type of authentication requires libraries to provide lists of the IP addresses associated with their institution to each of the publishers from which they purchase subscriptions. The IP addresses associated with an institution change over time, requiring continual updates to be distributed to each publisher. Pressures related to the shortage of IPv4 addresses and the transition to IPv6 further complicate the problem. In most cases, academic institutions with large networks will use generic IP addresses internally, which are translated to an external global IP address for access to resources beyond the local network. This process of network address translation (NAT) can also be problematic for cases when only specific departments subscribe to a resource but are not differentiated in the IP address seen by the publisher.

From the perspective of those accessing the resources, IP authentication is completely transparent when access originates from the campus network. Since the publisher recognizes the IP address of the user, no additional steps are required to view or download the resource. IP authentication works well at preserving the privacy of those accessing restricted resources. No specific information regarding the identity of the researcher is

IN THIS ISSUE

People in the News
PAGE 5

Smart Libraries Q&A
PAGE 6

passed into the publisher's infrastructure other than an institutional IP address.

The major failing of IP authentication relates to the reality that an increasing portion of access by people associated with an institution and authorized to access resources originate from IP addresses outside of its network. The students, faculty, and staff members of an institution expect to access library resources from off-campus locations that will not carry the institutional IP addresses. This problem is exacerbated by the skyrocketing use of mobile devices that may not use an institutional IP address even if used on campus.

Problems with IP Authentication

The problem of providing access to electronic resources is even more difficult for other types of libraries. Public libraries often acquire electronic databases and other content resources that likewise need to be restricted to their direct patrons. Since the constituency of authorized patrons come from diverse and unspecified IP addresses, the scheme of IP authentication does not work well at all for public libraries. It is common for public libraries to provide a login page where their patrons can enter their library card number or PIN to access resources. For programs offering statewide access to resources it is possible to use geolocation services to identify persons accessing resources within the authorized service area. Geolocation techniques offer a relatively low degree of accuracy, though they are generally within the tolerances accepted by publishers.

IP authentication represents a significant burden both for the providers of proprietary resources and for libraries. Each publisher must maintain a registry of the IP addresses associated with each of its customers and continuously update that registry. Libraries likewise must maintain a list of the valid IP addresses associated with their campus, which then need to be transmitted to each vendor when initiating a subscription and updated as the institutional network inevitably changes. For institutions that maintain separate subscriptions for specific schools or departments, the corresponding IP address sets must also be managed and distributed.

RedLink

The effort involved in IP authentication is substantial for both libraries and publishers. A company called RedLink has developed a service to assist libraries and publishers with managing IP addresses and other access credentials. The RedLink Network is a free service that enables libraries to provide their IP

addresses in one place, which is then available to all their content vendors. Publishers can then retrieve IP addresses from the RedLink Network for all participating customers instead of receiving them individually. In addition to its free services, RedLink sells products, such as the Library Dashboard and Publisher Dashboard, that offer statistical reporting tools and other features.

Proxy Services

The main technical approach that libraries have implemented to provide access to restricted resources for users not associated with institutional addresses involves the use of proxy services. A proxy service operates by performing some type of authentication for users outside the institutional network, and once authenticated, it conducts the resource request through an authorized IP address. From the publisher perspective, access continues to rely on IP authentication requiring no additional effort. Proxy servers can use a variety of mechanisms to validate users, such as a SIP request to the library's integrated library system, an internal database of users and passwords, or an institutional authentication service.

Access to a proxy service usually involves using a modified form of a URL to access restricted resources. The modified URL prepends the resource URL onto

the base URL of the proxy server. For example, the url *www.restrictedresource.com* would be re-written as *proxy.myinstitution.edu?url=www.restrictedresource.com*. Once the session is initiated, the proxy service will rewrite all links displayed from that resource to append its own URL so that they remain valid and authorized. To access resources via a proxy server, patrons usually need to access it through a catalog, discovery service, or by finding aid that provides the link in the modified form. Users accessing restricted resources via search engines or through bookmarked URLs from off-campus addresses will encounter the paywall.

The maintenance of a proxy server represents significant effort for a library since the URL for each resource must be registered in the proxy server, and all references to that resource in the library's environment must be adjusted.

Most proxy servers record each transaction they process in a log file. These logs can be processed to produce statistics on the resources accessed via the proxy server. Some universities even channel on-campus access through their proxy server so that they will be included within those statistics. It should be noted that proxy server logs represent only a subset of overall access since it does not include many on-campus users that access resources outside the library-provided interface or those

IP authentication represents a significant burden both for the providers of proprietary resources and for libraries.

that access resources from off campus using other forms of authentication.

A key weakness of proxy servers lies in the possibility of uncontrolled access to all remote resources if a single set of authentication credentials becomes compromised. Any individual gaining such access could perform wholesale downloads of restricted content from any of the providers available through institutional subscriptions. Most publishers monitor their services for such occurrences and will quickly disable access to that entire institution until the proxy issue is resolved.

EZproxy

The proxy services most widely used in libraries is EZproxy, an OCLC product. It was originally developed in about 1999 by Chris Zagar and sold through his company called Useful Utilities. OCLC acquired EZproxy from Useful Utilities in January 2008 (see the March 2008 issue of *Smart Libraries Newsletter*). Since its acquisition, OCLC has continued to develop EZproxy, which is now available in its sixth major version and is offered both as software for installation on a local Linux or Windows server or as a hosted service. EZproxy supports multiple authentication methods including LDAP (Lightweight Directory Access Protocol), CAS (Central Authentication Service), SIP2 (Standard Interchange Protocol Version 2.0), and Shibboleth.

Federated Authentication

Although IP authentication was initially a pragmatic solution, it has become increasingly problematic as the internet has evolved. But despite its limitations, this method persists as the dominant approach used to provide access to restricted scholarly resources even as more modern approaches have emerged.

The main alternatives to IP authentication rely on some type of federated identity management. Almost all educational institutions today have some type of centralized authentication service available used to provide secure access to all technology-based services. Rather than each application maintaining its own login scheme, most can instead rely on an external authentication service. These services can be based on Active Directory, LDAP, Kerberos, or other technology with a mechanism to validate the credentials of a user. Most institutional networks also offer a single sign-on capability so that once the user has performed the login sequence successfully for one application, access to other applications is granted without having to log in again. Some of the common single sign-on protocols include CAS, SAML (Security Assertion Markup Language), and Kerberos.

While single sign-on implementations work within a given institutional network, the problem of providing access

to restricted scholarly resources extends beyond that domain. Federated authentication has emerged as the main architecture able to solve this problem. The basic idea is that each institutional network implements its own scheme to authenticate its users, and access to services in another domain is allowed based on the trust between domains. A service does not need to know the identity of an individual making a request from an external domain. It only requires a mechanism indicating that the user was definitively authenticated within their home domain and that there is a previously established trust relationship among the domains that comprise the federation. In some cases, generic attributes are passed across domains to inform the authorization of resources, ideally without revealing personally identifiable information.

Several federated authentication services have been implemented, including

- **Shibboleth**, an Internet2 initiative initially launched in about 2000, which has steadily gained adoption. All of the components involved in Shibboleth are available as open source software, and it is designed to protect the privacy of users as they access resources across domains. Shibboleth is based on SAML and includes Identity Provider, Service Provider, and metadata aggregation components.
- **OpenAthens**, which is a single sign-on service based on Shibboleth and SAML. OpenAthens is offered by the non-profit organization Eduserv based in the United Kingdom. OpenAthens has been adopted by over 2,000 organizations, including many higher education institutions in the UK and internationally as well as by healthcare and research organizations, such as the National Health Service (NHS). In October 2017, Eduserv launched OpenAthens Cloud, which provides a less complex way for content providers to enable access to their resources compared to the locally hosted option.

RA21: Resource Access for the 21st Century

Providing access to remote restricted resources remains an unsettled issue. No single service or architecture has gained universal adoption, and basic IP authentication remains widely used despite its problems and limitations. Both publishers and libraries have strong interests in finding solutions that are technically sound, have a low level of difficulty and expense to implement, and that ensure privacy of access.

The International Association of Scientific, Technical, and Medical Publishers (STM) and National Information Standards Organization (NISO) have created a new initiative called RA21: Resource Access for the 21st Century. Launched in 2016, RA21 aims to solve the problems associated with

providing selective access to information resources on the web and to finally end the dependency on IP authentication. The initiative will work toward defining recommended practices, taking advantage of relevant standards and protocols, and will not define a specific technical solution. Principles for the initiative surround open solutions that avoid proprietary software or protocols, that can be implemented with a low threshold of difficulty, and that are neutral relative to any technology or content vendors.

Three pilot projects are currently underway, two in the academic sector and a third in the business environment. Each project is based on a different implementation of SAML to achieve federated identity management.

- The Corporate Pilot will validate the use of SAML technologies among pharmaceutical companies affiliated with the Pharma Documentation Ring. (see <https://ra21.org/index.php/pilot-programs/universal-resource-access-ura-pilot/>)
- Academic pilots:
 - Privacy Preserving Persistent WAYF is based on Shibboleth but incorporates additional information, such as the email domain into metadata exchanged across the federation. This additional information is termed WAYF (where are you from) hints (see <https://ra21.org/index.php/pilot-programs/p3-wayf-pilot/>).
 - WAYF Cloud project aims to validate the use of a cloud service for the exchange of data among publisher platforms (see <https://ra21.org/index.php/pilot-programs/wayf-cloud-pilot/>).

The pilot phase of the RA21 initiative is expected to run through early 2018, which will be followed by a possible publication of a NISO recommended practice that could foster future implementations.

For more information, see <https://ra21.org>.

Library Technology Vendors and ISO Certifications

Since libraries depend on external providers for critical aspects of their technical infrastructure, they require reassurance that those providers can deliver their services according to the highest possible standards. Individual customers of those businesses usually do not have the time or expertise to perform audits to verify performance in each area of concern. Instead, service providers apply for certifications that they comply to the requirements established by the International Standards Organization (ISO). ISO has developed standard

and measurable requirements in many different areas of technology and organizational practices.

Multiple vendors in the library automation industry have achieved conformance to ISO certifications in relevant areas of concern. Only some of these certifications are announced publicly, and there is not a comprehensive source detailing them. Some of the certifications that can be gleaned from public announcements and company websites include the following:

Basic IP authentication remains widely used despite its problems and limitations.

- **ISO 27018:** Protection of personally identifying information in a cloud computing environment (see <https://www.iso.org/standard/61498.html>).
 - Ex Libris (announced May 2016)
 - SirsiDynix (announced May 2016)
- **ISO 27001:** Information security management systems (see <https://www.iso.org/isoiec-27001-information-security.html>).
 - ODILO (awarded March 2016)
 - Ex Libris (announced April 2013)
 - OCLC (announced February 2012)
 - SirsiDynix (announced June 2015)
 - Innovative Interfaces (announced March 2017)
 - PTFS Europe
 - Axiell
- **ISO 22301:** Business continuity management and disaster preparedness (see <https://www.iso.org/standard/50038.html>).
 - ODILO (awarded July 2017)
 - Ex Libris (announced October 2017)
- **ISO 14721:** OAIS model for archiving and preservation systems (see <https://www.iso.org/standard/57284.html>).
 - ODILO (awarded July 2017)
 - Preservica
- **ISO 9001:** Quality management (see <https://www.iso.org/iso-9001-quality-management.html>).
 - ODILO (awarded October 2010)
 - OCLC
 - SirsiDynix (announced July 2015)
 - VTLIS (announced October 1997)

Absence from these lists should not be interpreted that other vendors have not achieved the relevant ISO certification. Libraries should ask their vendors regarding compliance in areas of interest. Many libraries will ask vendors to indicate specific ISO certifications on an RFP or RFI procurement process.

It should also be noted that vendors without ISO certifications can provide services at the same level of reliability, security, or quality as companies that have been officially recognized as compliant with ISO standards. The process for

certification can be extensive and expensive, which may be beyond the reach of smaller companies.

Resource Sharing Services: A Duopoly

Libraries have a long tradition of implementing technologies to help improve their services to their clientele through resource sharing. These services enable libraries to provide access to an expanded universe of materials available through peer institutions in addition to those they directly own. The number of organizations offering these services has consolidated, with OCLC and Auto-Graphics left as the primary providers. The open source FulFILLment software is also available but has seen limited adoption.

OCLC has been involved in resource sharing since 1979 when it introduced its interlibrary loan (ILL) service. The organization was created in 1967 to provide cooperative cataloging services. During the early phases of its history, OCLC provided its services via a private telecommunications network that it created to connect each of the libraries using its services to its data center in Dublin, OH. The organization gradually shifted from its private network to the internet in the late 1990s to early 2000s. OCLC's ILL system has evolved through multiple generations of technologies and interfaces, culminating in the current WorldShare ILL offering.

In addition to WorldShare ILL, OCLC has more recently developed a new product called Tipasa, which officially launched in January 2017. It manages the workflows of an ILL office to process borrowing and lending requests and communicates with WorldShare ILL and other services. Tipasa was developed on top of the WorldShare platform, not only taking advantage of its technical infrastructure but also providing a familiar interface for those libraries using WorldShare ILL. Tipasa was designed as a forward migration for the Windows-based ILLiad product, developed by Atlas Systems and exclusively distributed by OCLC.

The resource sharing sector has essentially consolidated into a competitive environment consisting of two major players—OCLC and Auto-Graphics.

In January 2017, OCLC acquired Relais International, which offered a peer-to-peer resource sharing platform. Relais D2D allows libraries within a consortium to enable mediated or unmediated requests of materials. The Relais ILL module provides the capability to send requests to external resource sharing systems, including OCLC WorldShare ILL.

Auto-Graphics has been in business since 1950. Originally a typesetting company, Auto-Graphics shifted to library services by about 1970. The company has offered an internet-based ILL management system since 1994, the precursor of its current SHAREit product. This product provides a union catalog, either through a physical aggregated database derived from catalog data from the participating libraries or through a virtual union catalog created dynamically via Z39.50. Its interlibrary request and fulfillment support standard industry protocols including NCIP and ISO ILL. It can connect with any major

external service that also supports those protocols. SHAREit has been implemented by many large-scale projects to provide ILL systems. In October 2017, Auto-Graphics hosted a conference for the users of its SHAREit service.

The resource sharing sector has essentially consolidated into a competitive environment consisting of two major players—OCLC and Auto-Graphics. OCLC provides services globally while Auto-Graphics operates primarily in the United States and Canada.

In its announcements, OCLC characterizes Tipasa as the “first cloud-based interlibrary loan management system,”¹ but that assertion is based on narrow definitions of cloud technology. It should be noted that Tipasa is a relatively new product and that long standing services from Auto-Graphics offer ILL transaction management and resource sharing based on internet connectivity and web-based interfaces.

1. OCLC, “OCLC Introduces ‘Tipasa’ Interlibrary Loan Management System,” press release, January 17, 2017, <https://www.oclc.org/en/news/releases/2017/201701dublin.html>.

People in the News

Infor has promoted **Jean-François (JF) Piat** to the role of General Manager for its Library Division, stepping in to the role vacated by Ann Melaerts. In his 13-year tenure with Infor, Piat has lead the Infor library division in France for 10 years

and has been acting General Manager of the Infor Library Division since August 2017. His sales career at Infor includes establishing V-smart as a leading system in France, including implementations in major public libraries such as the City

of Paris, Strasbourg, and Montpellier as well as the academic institutions of Bibliothèques Ste Geneviève and INHA. Infor representatives report that Piat embraces a customer-focused management style and is looking forward to building Infor's Library Division in all regions.

Tech Logic Inc. reports the death of its co-founder **Mark Frich** at age 66. Now a wholly owned subsidiary of The Library Corporation, Tech Logic was founded by Frich and Bob Rohlf

in 1997. According to the Tech Logic announcement, Frich, with an extensive background in engineering, designed the world's first automated material handling system for libraries. The Library Corporation acquired Tech Logic in 2005. Frich made significant contributions to the library community through the design of AMH equipment, was awarded a dozen patents for related inventions, and founded a company that has served libraries for over 20 years.

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 is the possible impact of artificial intelligence on library automation if it is to be applied?

Artificial intelligence has gained impressive capabilities in many aspects of society. This area of computer science continually makes new breakthroughs, which, for better or worse, introduce technologies able to perform activities of increasing complexity previously possibly only through human intelligence. The realm of libraries has already seen some impact from this type of technology and should expect even more in the future.

The key concept of artificial intelligence surrounds the ability for computers to perform tasks that traditionally rely on human intelligence. This concept goes beyond the automation of routine tasks. A bank ATM, for example, performs actions that are usually performed by a human teller simply by presenting menus of available services and responding according to pre-defined scripts. Robotics can likewise substitute for human labor in performing mechanical or logistical tasks without necessarily imitating human thought. Artificial intelligence goes beyond rote tasks to complex problem solving or other activities that take on characteristics of human cognition.

Computer technology, including artificial intelligence, has already reshaped the nature of library reference services. In former times, the library reference desk was considered the definitive source for factual questions. Resources such as Wikipedia as well as Google and other search engines able to provide instant access to authoritative and non-authoritative documents have siphoned almost all factual questions away

from library reference desks. The latest wave of technologies layers advanced technologies from the realm of artificial intelligence into the mix. Digital assistant services—such as Google Home, Amazon's Alexa, and Apple's Siri—take advantage of voice recognition, natural language processing, and machine learning to tap into information available on the web and proprietary data repositories to respond to ever more complex requests. While these services abound with concerns related to privacy, they nevertheless have further solidified a technical substitute to the library reference desk for routine fact-based questions. Persons with sensitive issues would hopefully seek out a librarian or other professional committed to safeguarding privacy.

Diverting routine factual questions from library reference services isn't necessarily a major loss. Many libraries instead channel their efforts into more in-depth information services for their clientele and other areas of strategic involvement with the communities they serve. Many librarians spend more time providing expert assistance with research projects, conducting bibliographic instruction or research methods sessions, or other activities beyond the reach of computerized assistants.

A strain of artificial intelligence that also has made inroads into the library domain is machine learning. A crucial aspect of library work relates to the ability to organize, describe, and provide access to large bodies of content. Machine learning describes a type of artificial intelligence where computers refine or perform new tasks based on processing a body of data. Traditional computer programming follows an algorithm to process and analyze data. With machine learning, the computer changes its behavior based on the data.

Library discovery services represent an interesting use case of machine learning. The common approach for discovery services today includes the creation of massive central indexes created through the processing citation or full text, representing some approximation of the totality of the body of scholarly

communications. Publishers and aggregators provide copies of the materials in their content products to discovery service providers that then ingest them into their central index. These discovery services can then be searched by library users to gain access to the original materials available on the publisher's delivery platform. This brute force indexing of citations does not fall into the realm of artificial intelligence since it is performing a routine task—albeit at large scale—taking advantage of standard search and retrieval technologies.

A new product from Yewno provides one example of using machine learning to support the search and exploration of large bodies of documents based on related concepts rather than keyword matching and relevancy ordering. Yewno has developed a process based on machine learning to not just index keywords, but to identify and extract concepts from documents. It is not merely finding matches of words in a text to a pre-defined ontology, but rather identifying concepts based on their semantic context within a document and linking those concepts to occurrences throughout a corpus of material. (*Smart Libraries Newsletter* featured Yewno in its April 2017 issue).

It is likely that artificial intelligence will make its way into many other aspects of traditional library activities over time. In some areas, such as broad discovery, we can expect artificial intelligence to result in new tools and interfaces with capabilities not previously possible with traditional search and retrieval technologies. I anticipate that various aspects of artificial intelligence can be tapped to improve the accessibility

of large-scale digital collections. Machine learning can power automated video description systems to facilitate the creation of search and retrieval systems for large digital video collections with more sophistication and at a scale not affordable solely through human processing.

Artificial intelligence should not always be seen as a threat to skilled workers. Tools based on artificial intelligence can also supplement the work of librarians. I would anticipate that new tools will be developed to create metadata to describe library resources. I see these kind of tools not as replacing the role of catalogers but as a means to amplify their work. Automated tools may be able to make a first pass at resource description; however, in most cases this work will require expert human intervention to ensure expectations related to following appropriate standards or neutrality and objectivity. I believe that it will be quite some time until libraries will be able to rely on entirely mechanical processes to create high-quality metadata to describe their core collections.

Artificial intelligence poses both a threat and opportunity for libraries. When approached proactively, these tools and technologies can be explored and exploited to facilitate library work and strengthen their position. On the other hand, some may see artificial intelligence as a way to eliminate or bypass libraries in providing information services. It will be important for libraries to continually assess each aspect of their work and shape services in ways that provide value beyond what might be delivered instead through technologies driven through artificial intelligence.

Questions or suggestions
for topics in future issues?



Contact Samantha Imburgia at
simburgia@ala.org



Smart Libraries Newsletter
American Library Association
50 East Huron Street
Chicago, IL 60611-2795 USA
Address Service Requested

NON PROFIT
US POSTAGE
PAID
PERMIT 4
HANOVER, PA

December 2017 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.

Production and design by the American Library Association
Production Services Unit.

Smart Libraries Newsletter is published monthly by ALA TechSource, a publishing imprint of the American Library Association.

alatechsource.org

Copyright © American Library Association 2017. All rights reserved.