Licensing Privacy Project

# Electronic Resource Authentication and Authorization

Sunshine Carter
Cody Hanson

June 2022

## Project Statement

The Licensing Privacy Project[1] seeks to use the power of library licensing agreements to effect change in third-party platform practices in order to bring them into alignment with library values of privacy, confidentiality, and respect for user control over their own data. It reflects an identified Pathway for Action from the IMLS-supported *National Web Privacy Forum*.[2] The goal is to develop model license language on user privacy that would support libraries in advocating for user privacy when contracting for services and content. By ensuring that user privacy is contractually protected in licensing agreements, service contracts, etc., libraries would be able to hold platforms accountable for their data practices.

## Funding and License Statements

This white paper was written by Sunshine Carter and Cody Hanson as a component of the Licensing Privacy Project led by Lisa Janicke Hinchliffe. This project was made possible in part by The Andrew W. Mellon Foundation.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

## Acknowledgements

We would like to thank the following people who graciously donated their time and energy in reviewing and providing feedback on a draft version of the white paper:

- Michael Berkowski
- Shaan Hamilton
- Megan Kocher
- Elizabeth Temple
- Stacie Traill

[1] https://publish.illinois.edu/licensingprivacy/
[2] https://scholarworks.montana.edu/xmlui/handle/1/15445

# Contents

# Introduction

Despite significant advances in the movement toward open access, the large majority of the scholarly literature and related tools and services that libraries provide to their users must be licensed from third parties. These license agreements obligate libraries to ensure that only those users stipulated in their license agreements can access these resources. If, for example, the University of Minnesota licenses a third-party database for use by current faculty, current staff, and currently enrolled students, we may allow use by these people, and only these people. To do this, we must be able to identify the user attempting to access the electronic resource, and then to determine if they fall into one of these allowed groups. These two steps are commonly known as authentication (identifying the user) and authorization (determining that user's access permissions).

In this paper, the authors provide an overview of common authentication and authorization mechanisms used in academic libraries today (these things change quickly, and so this paper is necessarily a snapshot in time). This paper is intended to provide academic library staff who do not regularly work with these technologies with a basic understanding of terms and concepts so that they are better positioned to contribute to conversations with their colleagues, and to inform users about how their information is used.

## Authentication and authorization

It was not always typical for a user to have to log in to their phone, a handful of different websites per day, or even to their computer itself. In the 1980s or 1990s a home computer user, if they thought about security at all, likely assumed that their computer was sufficiently protected by virtue of being inside their home.

Today, we are accustomed to logging in to many different devices and services on a regular basis. In doing so, we are making ourselves known to a computer system, either the one we are holding or typing on or the one delivering content to our web browser or app. We are connecting to the computer's stored knowledge of who we are: our shopping history, our saved searches, our friend list, or our email account(s). We are asserting our identity (typically in the form of a username or email address) and then proving it by demonstrating that we know a secret only we should know (typically a password). This workflow is often augmented by a second check (two-factor authentication, also known as 2FA) that demonstrates you have possession of something only you should have.

Logging into a system typically includes both authentication and authorization. Authentication is completed when the user successfully enters a username and password (and perhaps a 2FA code). This confirms that the user is authentically who they say they are. The system can then open up to the user those items or services they have permission to use. This is authorization.

For example, Cody can log in to our library's integrated library system and, once authenticated, can view a handful of reports in that system. He cannot, however, update catalog records. He lacks the appropriate permissions to do so. He is not authorized.

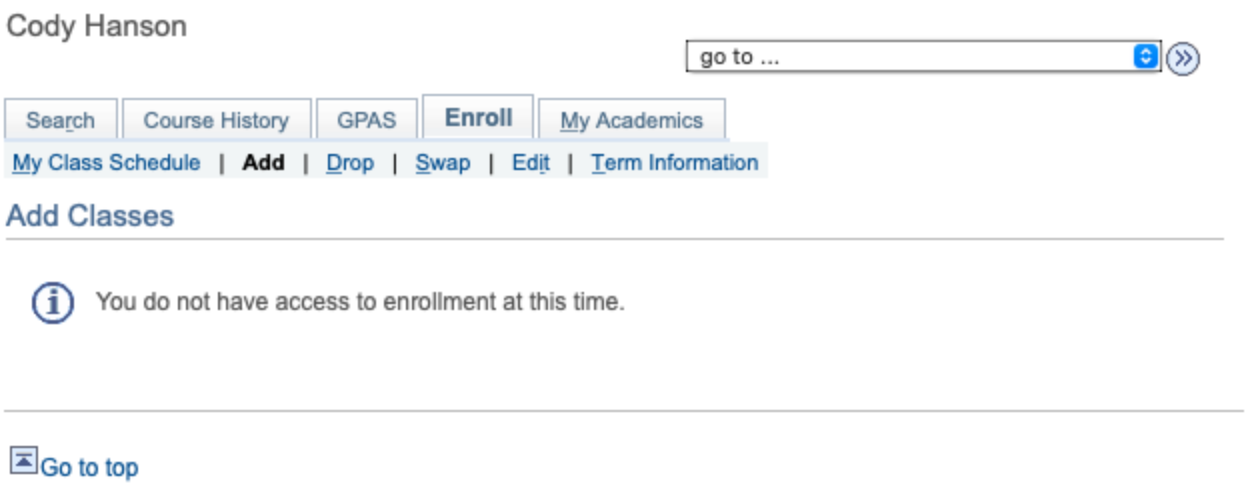


*Fig. 1 - Authenticated, but not authorized for enrollment*

At large institutions, the separation between authentication and authorization can be more readily apparent. If your college or university has a single sign-on (SSO) infrastructure, you likely encounter the same login screen over and over again when accessing different systems and services. Provided you remain enrolled and/or employed, you can always log in through that screen, but that doesn't mean you can always access the system behind it. For example, as seen above, as a staff member at a university with SSO, Cody can authenticate into the student course registration system but, because he is not enrolled as a student, he is not authorized to access the system's registration functions. Authentication, the determination of who he is, is handled by our SSO system, which passes that information along to the registration system to determine authorization, based on registration permissions.

## Electronic resource authentication and authorization

The rest of this paper is devoted to describing some of the ways authentication and authorization can be implemented for library electronic resources. Many libraries must use more than one method and the choices they have will be determined by both what is supported by the providers of electronic resources as well as what is supported by their campus. Depending on the diversity of a library's collection, it is likely a library will be required to implement more than one of these methods. Some larger and more sophisticated content providers may require direct user authentication via SAML (Security Assertion Markup Language, the technology behind what many in higher ed know as Shibboleth). Some smaller content providers still require the use of the same shared Username and Password for all users from a given institution. Many content providers support a handful of different mechanisms, recognizing there is a broad range of technical capacity among their library customers.

This paper is part of a larger project investigating privacy issues in library electronic resource licensing and access, *Licensing Privacy*[3]. While the paper will not go into great detail about the privacy implications of various authentication and authorization methods, it is worth noting that authentication and authorization are both premised on identification of a user. Some authentication and authorization

[3] https://publish.illinois.edu/licensingprivacy/

strategies are more privacy-preserving than others; however, in most cases one or more of the library staff, campus network administrators, or the resource provider may be able to identify users and the resources they access. It is increasingly difficult, if not impossible, for an academic library to provide truly private, anonymous access to licensed electronic resources.

The most prevalent mechanisms used today by content providers and libraries to provide access to subscribed electronic content are described below. For clarity, we will consider authorized users of an academic institution to be current students, faculty, and staff of the institution (part-time or full-time), regardless of location, and walk-in users who are physically present at the institution.[4]

## Username/password

The simplest access mechanism is username and password (U/P). Usernames are assigned by the content provider, and anyone with knowledge of the username and password can gain access. U/P access may be for a single user or shared by multiple users. Passwords, particularly shared passwords, should be updated regularly to limit unauthorized use.

Some content providers may also allow libraries to post U/P credentials behind other authentication mechanisms to share usernames and passwords with an authorized set of users (for example, current students, faculty and staff) rather than post them for the entire world to see. These content providers may also allow for credentials to be shared directly with a user at a service desk. Other content providers require each user to have their own individual username and password to gain access, created through a self-registration process or with library assistance. Individual personal usernames and passwords associated with a specific named person are not private; they may be known to the library, the content provider, or both.

Many content providers also offer users the ability to register for an individual account (separate from the library-established authentication mechanism) to access additional functionality such as saved searches and results, notifications, and organizing. These opt-in features may further reduce user privacy.

## Internet Protocol (IP) based access

Most users of academic library electronic resources are authorized and authenticated through internet protocol (IP) address recognition. IP addresses are a string of characters (for example, 134.84.192.101) that resource providers can use to identify the network or Internet Service Provider (ISP) a device is using to connect to the Internet. Libraries give content providers a list of IP addresses and/or register them with an IP registry service (such as IP Registry[5]). Devices connecting to the resource using one of these IP addresses are granted access to subscribed content.

Libraries provide the approved IP address ranges at the point of setting up the electronic resource or when IP addresses change. Changes to institutional IP address ranges can pose problems for IP-based access. Changes require notification to the content provider or third-party registry service and,

[4] Liblicense Model License Agreement (“LMLA”). November 2014 revision. http://liblicense.crl.edu/licensing-information/model-license/
[5] https://theipregistry.org/

depending on the number of content providers. This can be onerous. ISPs and campus networks are in the process of transitioning from IP version 4 (IPv4) to IP version 6 (IPv6). IPv6 addresses are more complicated to manage, and many content providers cannot yet authorize users based on IPv6 addresses.

### Institutional networks

Obtaining an IP address for an institutional network is often easy for current students, faculty or staff. They simply connect their device to the campus's wired or wireless network. Depending on how a library sets up electronic resources with content providers, this may be sufficient to grant access to library electronic resources. It may be more complicated if there are other authorization mechanisms in place or if the number of simultaneous users is limited.

On-campus authentication and authorization to electronic resources may seem seamless, but off-campus use of the same electronic resources typically requires users to log into an institutional network remotely using VPN, or employ some other mechanism (for example, EZproxy, single sign-on, Google CASA).

#### VPN

Electronic resources are often accessible through virtual private networks, or VPNs. Authenticating and connecting through a VPN creates a secure and encrypted connection to a private campus network, useful for making remote users appear as though they are operating from within the campus network. A VPN will also have associated IP addresses (one or many, depending on the size of the institution). If these IP addresses are among those given to content providers, then users who are connected to the VPN can access electronic resources.

#### eduroam

Given the importance of eduroam as a component of academic institutional networks, it is useful to understand how eduroam enables IP authorization. eduroam (education roaming) was created to help students, faculty, and staff connect to the Internet while traveling to other participating institutions. eduroam allows users to log in to another institution's WiFi using credentials from their home institution.[6] Some institutions recommend their affiliates use eduroam to connect to wi-fi even on their home campus. eduroam is currently available at thousands of institutions in 100+ countries.

Devices connected via eduroam obtain an IP address from whatever network they are connected to (home or visiting institution). Given this, a library must understand how the campus sets up the eduroam service in determining whether the eduroam IP addresses are provided to content providers for authorized access. At the University of Minnesota Libraries, we do not provide our eduroam IP addresses to content providers because we do not consider being on eduroam the equivalent of being a walk-in user in our libraries.

### Proxies

A proxy alters the IP address that a user's device presents to a content provider by redirecting web traffic to that provider through a server that the library controls. This makes the user appear to be

[6] https://eduroam.org/

operating from within the campus network even if their device does not have an institutional IP address. The proxy server's IP address is among those in the institutional IP range. Libraries use proxy servers to facilitate access to electronic resources off-campus, because off-campus users are outside of the institutional IP address ranges.

EZproxy[7] is a common proxy used by libraries of all types. To use EZproxy, a library must set up an EZproxy authentication/authorization method that works for its established infrastructure, tools, and mechanisms. EZproxy works with a wide range of user and authentication systems used by libraries and higher education. It is possible to use a simple list of Username and Password combinations to authenticate/authorize users solely with EZproxy; but, this method is less secure and not current best practice.

Once a user is authenticated, EZproxy modifies the user's destination URL to direct their web traffic through the library's proxy server and presents the EZproxy server's IP address to the content provider. The content provider recognizes the EZproxy IP address as an authorized IP address – coming from an institutional subscriber – and allows the user to access the content.

Unfortunately, not all electronic resources can be accessed through a proxy and the increasing complexity of content provider websites is making proxying more difficult to configure and more prone to breaking with each site update. Streaming media and corporate business resources in particular are frequently incompatible with proxy authenticated access and require that users either log in directly with a Username and Password or connect to the institutional network through VPN.

## SAML (Security Assertion Markup Language)

Security Assertion Markup Language (SAML) is an open standard for exchanging information about users, authentication, and authorization between systems[8]. Strictly speaking, SAML is not an authentication or authorization method in and of itself; it is a shared language that allows for a variety of authentication systems to interoperate.

### Shibboleth

In higher education, one of the most common implementations of SAML is Shibboleth[9]. Shibboleth is a single sign-on technology that allows users to authenticate to many different systems with a single Username and Password (and where possible, 2FA). For example, at the University of Minnesota we use Shibboleth for single sign-on. As a result our users see the same log-in screen and use the same credentials to borrow from our library catalog, access our electronic resources (via EZproxy), and use their umn.edu Gmail account, the University's Canvas learning management system, and dozens of other applications.

Authentication and authorization with Shibboleth uses a model consisting of an Identity Provider (IdP) and a Service Provider (SP). Each site, service, or application that wishes to authenticate users via Shibboleth is a Service Provider. When a user initiates a login to the Service Provider, they are

[7] https://www.oclc.org/en/ezproxy.html
[8] https://wiki.oasis-open.org/security/FrontPage
[9] https://www.shibboleth.net/

redirected to the Identity Provider to authenticate, with a familiar log-in screen. Upon successful authentication, the Identity Provider redirects the user back to the Service Provider, sending along some SAML metadata. This metadata is then used by the Service Provider to determine authorization.

For the University of Minnesota, many of our alumni continue to use their umn.edu email address. However, with few exceptions, alumni are not allowed to access library electronic resources remotely. Shibboleth workflows for an alumni user of our Gmail and EZproxy systems function something like this:

> **Gmail**: An alumni user with an active umn.edu email account visits gmail.com. On a Gmail login screen they are prompted for their email address. Gmail recognizes the umn.edu domain on the email address, and redirects the user to the University of Minnesota Identity Provider for authentication. The user sees a familiar University of Minnesota log-in screen, enters their Username and Password, and then is redirected back to Gmail. The UMN IdP also sends along to Gmail SAML metadata with the user's Username, email address, and confirmation that the user is authorized to use UMN Gmail. Gmail uses this information to authorize the user into their email account. The user is authenticated and authorized for this service.
>
> **EZproxy**: An alumni user visits the University of Minnesota Libraries website and clicks a link for one of our licensed databases, all of which are accessed through EZproxy. The user is redirected to one of our EZproxy servers. EZproxy attempts to authenticate the user by redirecting them to the IdP. Here, again, the user sees the familiar UMN log-in screen. They are able to successfully authenticate with the IdP and are redirected back to EZproxy, the Service Provider. However, the metadata returned by the IdP to the SP indicate the user is not authorized to access library licensed resources. EZproxy then redirects the user to a page explaining the library's access policies. The user is authenticated but is not authorized for library resource access.

A number of library content providers are also able to act as Service Providers, authenticating users directly to their platforms via Shibboleth or similar SAML mechanism. This use case looks very much like the Gmail example above, with authorization information sent directly from the Identity Provider to a vendor Service Provider. The Seamless Access[10] (formerly RA21, Resource Access for the 21st Century) initiative developed by international standards bodies and publishers is an effort to standardize this type of direct authentication, including determining how much identifiable user information (if any) must be shared from the IdP to the vendor SP.

The complexity of Shibboleth, and its design as a single sign-on technology, means it is often implemented and managed by a central campus technology unit rather than the library. Successful use of campus Shibboleth for library systems will likely require cooperation from the unit managing the Identity Provider for the university.

[10] https://seamlessaccess.org/

### OpenAthens

OpenAthens[11] is a vended product that provides library-centric single sign-on for systems and licensed resources. OpenAthens can replace all or some of a library's proxy server, e-resource management tools, and single sign-on system. OpenAthens can use SAML to manage authentication and authorization for electronic resources.

## Google CASA

Google Inc. (now Alphabet Inc.) built a suite of tools to search and link to scholarly content. Google Scholar[12] indexes publisher content for discovery. Google Scholar Library Links direct Google Scholar users to full text copies of content (when available) through their institution's link resolver (for example, SFX, Alma, 360Link, etc.[13]). Released mid-2017, Google's Campus Activated Subscriber Access (CASA) attempts to improve access to authorized copies of scholarly literature when users are off-campus. Publishers implement CASA to allow content to be accessed through this Google service. When they do so, CASA is enabled for library users without any action on the part of a library.

CASA uses an encrypted token to identify a user's browser as authorized to access subscribed content from participating publishers.[14] The token is placed when a user accesses subscribed content from a campus network connection, while using a Google service or application. When the same academic user travels off-campus and uses the same browser, the token continues to confer authorization to access institutionally subscribed content for up to 30 days. Tokens are renewed when a user revisits a participating publisher from a campus network IP.

CASA originally required use of Google Scholar to get access to content from off-campus, but Universal CASA[15] (released mid-2018) eliminates that requirement and provides access to any participating publisher's content regardless of the starting point. This means that the user could go directly to a CASA-enabled publisher website and be authorized via the token.

Publishers and content providers continue to implement CASA/Universal CASA. Individual scholars can opt out of the CASA token placement; institutions cannot. The CASA token as a successful authentication mechanism depends on many factors such as browser cookie settings, token renewal, and campus visits (which have become less predictable due to the COVID pandemic and changing policies around remote work and study). Additionally, web browsers have more restrictive default security settings than in the past, and users are increasingly likely to further modify browser and computer settings to increase their privacy, thus making CASA tokens less likely to be retained.

---

[11] https://www.openathens.net/
[12] https://scholar.google.com/; https://scholar.google.com/intl/en/scholar/libraries.html
[13] https://exlibrisgroup.com/products/primo-discovery-service/sfx-link-resolver/; https://exlibrisgroup.com/products/alma-library-services-platform; https://exlibrisgroup.com/products/summon-library-discovery/360-link/
[14] The authors believe the mechanism for token establishment is (or is similar to) that described in US Patent 16/122,294. Location-based access to controlled access resources. A Acharya, S Yuan, AA Verstak - US Patent App. 16/122,294, 2020. https://patentimages.storage.googleapis.com/a3/40/e3/18b680a877f523/US20200052896A1.pdf
[15] https://www.highwirepress.com/technological-innovation/highwire-leads-industry-rollout-of-universal-casa-in-partnership-with-google-scholar/

## Conclusion

The paper provides a simple introduction to authentication and authorization of library resources. Academic libraries and institutions implement a variety of complex systems and tools to facilitate access to library electronic resources. Library staff need to understand the authentication and authorization landscape to maintain access to library resources, facilitate troubleshooting, and assess the impact of potential changes to systems. Though this area of academic librarianship will continue to evolve rapidly, it is our hope that this paper will empower readers with sufficient information to approach current and future conversations about electronic resource access with confidence.