Leveraging Federated Authentication to Simplify Access: Understanding Changes in Access Mechanisms to Online Content

John Felts and Todd Carpenter

Abstract

The COVID-19 pandemic is profoundly reshaping access to education and one of the areas impacted is access to electronic resources. While remote access has been a feature of our industry for decades, it was generally viewed as the exception. Now all of us have become remote users. This paper uses case studies to re-examine the remote access experiences of three libraries that navigated the shift to majority (or exclusive) remote access and how federated access has proven to save time and money for both small and large institutions alike. It also presents the genesis and ongoing evolution of the SeamlessAccess project, addresses privacy concerns related to accessing online content, and identifies potentially disruptive changes in browser technologies that are on the horizon.

Keywords: SeamlessAccess, federated authentication, user privacy, browser changes

Libraries in the Pandemic

In early March 2020 COVID-19 blindsided academic libraries. With little time to plan, we closed our library facilities to protect the safety

of our patrons and employees and moved to online services only, working from home. Years of curating online content while developing intuitive, user-friendly web presences and robust search interfaces helped us to meet the research needs of our users during this challenging time. However, many libraries discovered that there were problems with the technologies used to provide remote access to online resources. Virtual private network (VPN) servers buckled under the weight of a significant increase in traffic and libraries found that many users who seamlessly accessed resources while on-campus, struggled when attempting remote access via a proxy server. In this section, the access experiences of Boston College Libraries, the Dahlgren Memorial Library at Georgetown University Medical Center, and Kimbel Library at Coastal Carolina University are reviewed.

Emily Singley, the head librarian for systems & applications at Boston College, presented a case study of how the pandemic impacted access in Fall of 2020.[1] Prior to March 2020, she found that 97 percent of library access at Boston College was Internet Protocol (IP) authentication, of which the vast majority was users accessing directly from on-campus IP addresses. Only 17 percent of access required users to log in via the library proxy server. Federated authentication was only enabled for Elsevier's ScienceDirect and accounted for only three percent. When comparing resource usage information for March 2019 and March 2020, Ms. Singley found that e-resource usage fell by 45 percent when the pandemic hit, but ScienceDirect actually saw its usage increase 35 percent. She attributed much of this change to the library's support for federated authentication (the orange bar in Figure 1), which was much easier for students to use when accessing resources directly while remote. While overall IP authentication in general fell a little in March 2020 compared with March 2019, (the blue bar in Figure 1), federated authentication took off like rocket, growing from 3 percent to 57 percent of total ScienceDirect access.

Linda Van Keuren, assistant dean for resources and access management at Dahlgren Memorial Library, Georgetown University Medical Center, also shared her perspective. Dahlgren Library's administration
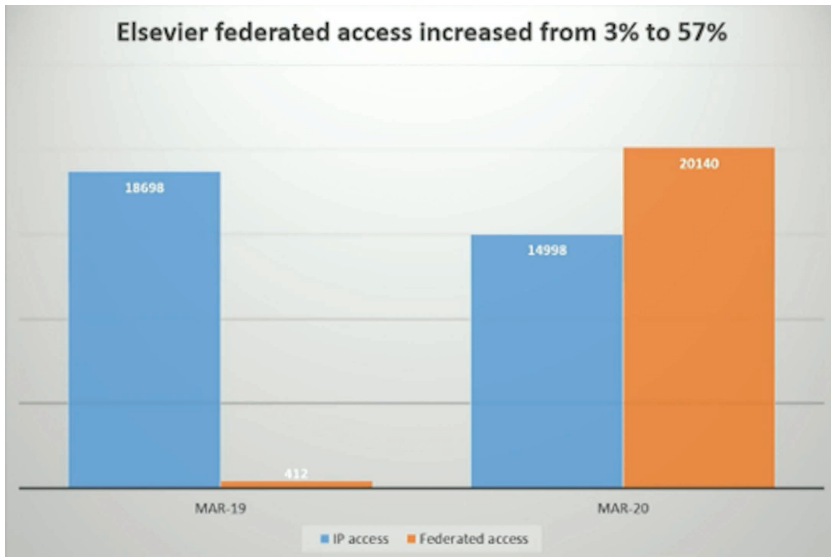
Figure 1. ScienceDirect access data for Boston College.

identified long-term goals for e-resource management and access in 2016, which were:

- Stronger security since user access and authentication to all platforms is centrally managed
- Access entitlement determined by user identity rather than location
- Ability to generate much richer resource usage data that includes user demographic information
- Ability to select and provide access to targeted users or departments at their university
- Provide a far better user experience that provides authentication at the point of discovery

These goals were all achieved by moving from IP access to federated authentication and when the pandemic began, the library was already positioned to provide a consistent, familiar user access experience to its user community.
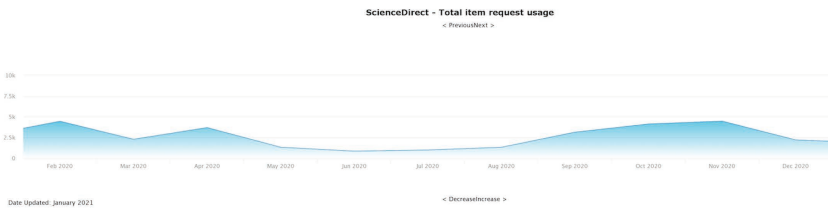
ScienceDirect - Total item request usage
< PreviousNext >

Date Updated: January 2021   < DecreaseIncrease >

Figure 2. ScienceDirect access data for CCU.

Lastly, Coastal Carolina University (CCU) did not have quite the challenges that many libraries experienced because its on-campus population is smaller than many universities. Therefore, roughly 75 percent of their student population had already been enculturated into accessing resources remotely via EZproxy since they live off-campus and outside of the CCU IP range. However, they did notice a considerable uptick in the number of support requests that were fielded in the first couple of months of the pandemic and consistently thereafter. These issues almost exclusively revolved around confusion with users who started their research from the open web, eventually made their way to a journal article residing behind a paywall, and then had no idea how to authenticate for access to content which they were rightly entitled.

But there was one exception to this scenario. As a result of CCU's participation in Elsevier's Federated Access-Only initiative, federated access to ScienceDirect using CCU's local Shibboleth software had already been configured. Therefore, no support calls related to ScienceDirect access were submitted and their annual usage patterns remained very consistent compared with years prior.

Problems with IP-Based Access

Consider that the entire IP filtering model is constructed on the assumption that an IP address reliably indicates a user's physical location. With proxy servers and VPN clients, this is simply no longer the case. To compound the issue with location-based access, this model assumes that a physical location can be relied on to indicate a legitimate, authorized user, which is demonstrably false. There has been a

considerable increase in off-campus access to library resources over the years from a multitude of devices. IP filtering has been obsolete for quite some time because it is concerned with <u>where</u> an anonymous user is, instead of <u>who</u> the user is.

IP authentication is also counterintuitive to the current research user experience because the user is forced to start from—or at some point circle back—through the institutional portal before they can access the online content to which they are rightly entitled. This is contradictory to the way discovery works. It degrades the overall user experience and needlessly complicates research. IP filtering is unsecure and easily exploited. In a presentation at the Society for Scholarly Publishing Conference in June 2017, Craig Griffin of Silverchair Information Systems suggested "Without IP filtering, Sci-Hub could not exist."[2]

Also consider that IP filtering is indiscriminate. If a user engages in downloading behavior that breaches the provider's license agreement, the proxy server connection will be blocked, cutting off access for *all* users. Library use of IP recognition was developed when off-site access to electronic resources was in its infancy and has changed very little since then. Meanwhile, IP addresses are no longer static, and devices are no longer stationary. IP based filtering is completely unsuited to the current internet environment. In fact, authentication is being further broken right now by changes browser providers are making to implement privacy preserving practices. IP-based tracking has been eliminated or obfuscated by browsers, and Apple has already implemented this feature for subscribers of their iCloud+ service. Browsers using the WebKit engine, so essentially Safari and any browser running on an iOS device, may exhibit this behavior. This paper, along with an FAQ published by SeamlessAccess further explores this development.[3]

Benefits of Federated Access

The preferred alternative to IP filtering is Federated Identity Management. It utilizes a more reliable, robust, Security Assertion

Markup Language (SAML) based infrastructure, typically using Shibboleth, and therefore does a better job of protecting user privacy since it passes selective information about an individual to service providers from their identity provider without giving out the user's credentials.

Also, when compared to EZproxy, ongoing maintenance of this environment is much simpler since the specific information necessary to establish a trusted connection lives in a federation accessed by identity provider and service provider alike. It also provides support for researchers who might be affiliated with numerous institutions or research organizations. Lastly, federated access offers a far better user experience in that provides authentication at the point of need.

Recent findings in Coastal Carolina University's migration from EZproxy to OpenAthens in August, 2021 supports the need for authentication at the point of need. CCU re-encoded, or "Athenized," all resource links in their A-Z lists and discovery platform and went live with OpenAthens on August 16, 2021. The infrastructure was in place and functional beginning on July 22, 2021, but it was hidden from users. In checking their OpenAthens usage dashboard, CCU noticed they were receiving inbound traffic and successful authenticated logins. There were 1,512 successfully authenticated sessions in this time
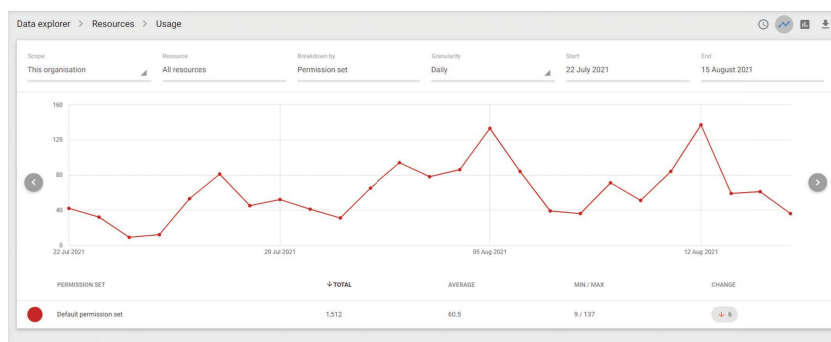

Figure 3. CCU's OpenAthens authenticated sessions: July 22 – August 15, 2021.

period with as many as 137 sessions in a day. The explanation for this occurrence is that users were accessing resources via the Wayfinder service on each vendor's platform after having started their journey from the open web, and were then intuitively able log in with no help or intervention from library staff.

User Experience

The user authentication experience can be inconsistent, confusing, and replete with jargon. Accordingly, SeamlessAccess is implementing a standard for federated authentication based on a single sign-on through the user's home institution. Regardless of where end users begin their research they will encounter consistent imagery, language, and login placement, along with a standardized identity provider discovery experience.

Once authenticated using their preferred affiliation credentials, the end user will not be required to sign in again across all SeamlessAccess-enabled sites.
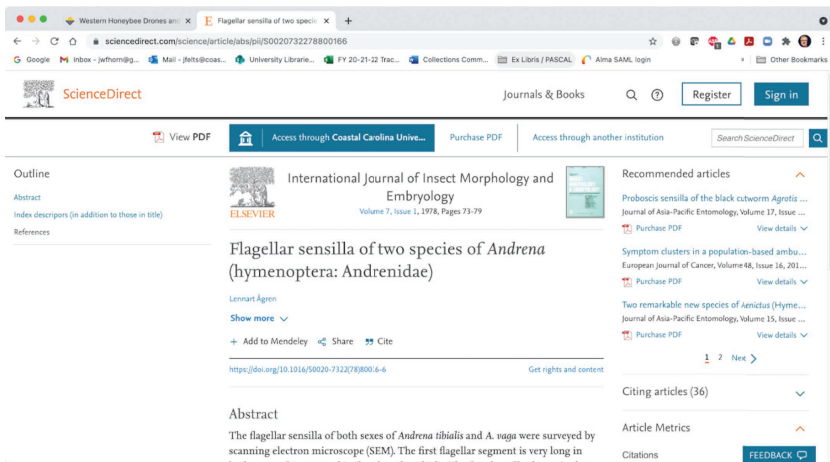

Figure 4. SeamlessAccess: consistent imagery, language, and placement.

SeamlessAccess

SeamlessAccess was created in July 2019 as a community-driven initiative to enable seamless access to information resources, scholarly collaboration tools, and shared research infrastructure. It builds on the existing federated authentication infrastructure that enables users to authenticate using their existing institutional credentials—a process widely known as Shibboleth or OpenAthens. Federated authentication is preferred to IP addresses as a method of authenticating users because it offers authentication at the point of discovery. Federated authentication does not route remote users through a proxied URL, supports personalization without requiring users to register yet another set of personal credentials with a vendor, and protects user privacy by enabling pseudonymous access. Pseudonymous access addresses one of the major concerns from libraries with respect to individual logins. Federated authentication is also more secure and requires less ongoing administration than IP address authentication.

SeamlessAccess improves access by addressing one of the major sources of friction in federated authentication, which is when a user is required to select their institution every time they login. The SeamlessAccess project is also developing important best practice recommendations to standardize and simplify access experiences for users. It currently has four founding organizations:

- National Information Standards Organization (NISO)
- Internet2: a U.S. research and education (R&E) network that operates the U.S. identity federation, InCommon, among many other activities
- GÉANT: a European research and education network that operates a service called eduGAIN that connects over sixty R&E identity federations around the world. GÉANT also supports related initiatives, such as FIM4L
- International Association of STM Publishers

SeamlessAccess grew out of a project called Resource Access in the 21st Century (RA21) that was initiated in 2016 to explore the challenge of remote access.[4] It involved stakeholders from the publishing, library, software, and identity communities and collected input from sixty organizations over the course of three years. RA21 revealed that federated authentication held the most promise for providing a robust, scalable solution for remote access to scholarly content and investigated barriers to implementation, developed best practices, and piloted technical approaches to simplifying access. RA21's conclusions were published as a final NISO Recommended Practice in June 2019.[5]

Information designed for a non-technical audience on federated authentication and SeamlessAccess is in the SeamlessAccess Learning Center.[6] It includes a series of short videos explaining how federated authentication works, how it protects user privacy, how SeamlessAccess improves the researcher experience, a terminology guide, and a library case study.

A growing number of global publishers have enabled SeamlessAccess, including Elsevier, Wiley, Taylor & Francis, and Springer Nature. Integrations are also underway with a number of the most popular publishing platforms, such as Literatum and Silverchair, which will rapidly increase the number of library vendors across the industry that support SeamlessAccess.

Ongoing Work

There are several important initiatives currently underway at SeamlessAccess.

Entity Categories

One of the major challenges that libraries face with federated authentication is configuring what personal data to share with vendors. To

simplify this process, SeamlessAccess has proposed two new categories called Anonymous Authorization and Pseudonymous Authorization that provide pre-determined sets of attributes.[7] The former is entirely anonymous and prevents any form of personalization. The latter allows publishers to recognize a returning user but only with a pseudonymous identifier, thereby supporting personalization. However, before these categories can be widely adopted in practice, additional work is required to think through the implications.

Contract Language

SeamlessAccess has a Contract Language Working Group that is currently investigating the development of contract language templates for library use based on the two new entity categories. This gives libraries a mechanism to ensure Attribute Release compliance. They have been examining existing contract language and determined, perhaps unsurprisingly, that there really is no consistency across vendors and that a lot of contract language in use does not reflect current technologies. The Working Group recently released a proposed set of use cases for the entity categories that will form the basis of contract language templates.[8]

Upcoming

SeamlessAccess is in the process of becoming a Stichting, which is a Dutch non-profit corporation. It will be a partnership among the core founding organizations (Geánt, Internet2, NISO, and STM), with the possibility for additional partners and expansion when appropriate. This will provide SeamlessAccess with the legal framework for managing the system and will serve as an entity for managing the collective budget, holding intellectual property, enforcing legal terms of service, ensuring appropriate use, and establishing contracts with vendors and support staff.

Prepare/Inform/Advocate

Librarians should become more familiar with the aforementioned browser issues since troubleshooting with users will be complicated. It is important to evaluate current access methodologies and understand how these changes may affect operations and the user experience. Libraries should be prepared for a series of changes in how authentication, authorization, and access controls are understood and implemented, and it is therefore critical to open lines of communication with university information technology (IT) departments.

It is unlikely that authentication will suddenly break in large numbers, but because of the complicated nature of access in general, communication will be equally complicated because it involves numerous stakeholders including the organization, library service providers, library users, campus IT departments, and other stakeholders. It is equally important to internally and externally advocate about these upcoming changes. We conclude with the following recommendations:

- Work with IT departments to help understand the specific data being shared and the decisions being made regarding those data.
- Work with colleagues who handle the licensing of e-resources and software to understand the effect that federated authentication has on existing contracts. As your library renews contracts, consider adopting language specific to federated authentication where necessary.
- Look to groups that are paying attention to and working on the big picture of these changes, such as SeamlessAccess, W3C, and NISO.
- Be ready to work with groups outside your organization (federations and other large cooperatives) to find solutions.

Contributor Notes

John Felts is Head of Information Technology and Collections, Coastal Carolina University, Conway, South Carolina.

Todd Carpenter is Executive Director, National Information Standards Organization, Baltimore, Maryland.

Notes

1 As reported in Library Connect, "Maintaining Access to Library Resources in a COVID-19 World," Elsevier, accessed November 15, 2022, https://www.elsevier.com/connect/library-connect/maintaining-access-to-library-resources-in-a-covid-19-world.

2 Jacob Wilcock, "Putting an End to Download-and-Go: The Website's Role in a Content Marketing Ecosystem," *Learned Publishing* 31 (2018): 173–79, https://www.atypon.com/wp-content/uploads/2018/08/Content-Marketing-Scholarly-Publishing_Learned-Publishing_Atypon.pdf.

3 "FAQ on Browser Privacy Changes and Library Resource Access," SeamlessAccess.org, accessed November 15, 2022, https://seamlessaccess.org/learning-center/browser-faq.

4 "Resource Access for the 21st Century," RA21, accessed November 15, 2022, https://ra21.org/.

5 NISO, "NISO RP-27–2019, Recommended Practices for Improved Access to Institutionally-Provided Information Resources: Results from the Resource Access in the 21st Century (RA21) Project," accessed November 15, 2022, http://www.niso.org/publications/rp-27-2019-ra21.

6 "Learning Center," SeamlessAccess.org, accessed November 15, 2022, https://seamlessaccess.org/learning-center/.

7 "Entity Category Use Case Scenarios," SeamlessAccess.org, accessed November 15, 2022, https://seamlessaccess.org/posts/2021-06-17-contract langusecases/.

8 "Federated Authentication Model License Agreement w/Commentary," SeamlessAccess.org, accessed November 15, 2022, https://docs.google.com/document/d/1scLrPQMDTtl4j7F7oqnQCPPQFBa2IB9P8dlJe2XYDUs/edit?usp=sharing.