2021

SeamlessAccess.org: Delivering a Simpler, Privacy-Preserving Access Experience

John Felts
Coastal Carolina University, jfelts@coastal.edu

Recommended Citation

Felts, J.W., Jr., Lloyd, T. (2020). SeamlessAccess.org: Delivering a Simpler, Privacy-Preserving Access Experience. Charleston Library Conference Proceedings, doi: https://doi.org/10.3998/mpub.12470905

SEAMLESS ACCESS: DELIVERING A SIMPLER, PRIVACY-PRESERVING ACCESS EXPERIENCE

John Felts, (jfelts@coastal.edu) Head of Information Technology and Collections, Coastal Carolina University
Tim Lloyd, (tim@liblynx.com) CEO, LibLynx

ABSTRACT

Managing access to subscribed services in an era of abundance is a major challenge for libraries. Users have come to expect a seamless, personalized experience on their mobile devices, but traditional approaches to access management force librarians to choose between the anonymous ease of onsite IP authentication or the access friction experienced by users authenticating via a proxy server or across multiple resources with Single Sign-On. Building on the work of the RA21 initiative[1], a recent NISO Recommended Practice on Improved Access to Institutionally Provided Information Resources, Seamless Access charts a way forward. It will enable libraries to provide seamless, privacy-preserving and one-click access to subscribed content from any device, any location, and from any starting point in the research process. Seamless Access builds on both RA21 and the NISO Recommended Practice and is currently in a beta phase implementation. But how is user and data privacy protected, how is access simplified, and how is Seamless Access helping libraries implement this service? This paper discusses how these concerns are being addressed by a consortium of industry partners including librarians, access providers, publishers, and standards organizations. It also discusses how Seamless Access will manage this service for publishers and libraries while continuing to improve this user experience, provide governance on data policy and privacy issues, and maintain core web services specific to this initiative.

WHAT IS SEAMLESS ACCESS?

Seamless Access is the operational successor to the Resource Access in the 21st Century project (RA21), and is a community-driven effort to enable seamless access to information resources, scholarly collaboration tools, and shared research infrastructures. To date, this initiative has four founding organizations: the National Information Standards Organization (NISO), GÉANT, Internet2, and the International Association of STM Publishers, and features a full-time implementation team that includes an experienced library technologist dedicated to library outreach. There's a governance committee with representatives from across stakeholder groups, an outreach committee that includes six institutional participants, and two cross-industry working groups. The Attribute Release Working Group is developing standards for attribute release in the context of library resources, and the Contract Language Working Group is tasked with developing contract language templates to give libraries a mechanism to ensure attribute release compliance.

WHY DO WE NEED SEAMLESS ACCESS?

[1] https://ra21.org/

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. IP recognition has been around since the 1970's, and 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. In fact, libraries are one of the few organizations that still uses IP filtering. To compound this issue, the location-based access model assumes that a physical location can be relied on to indicate a legitimate, authorized user which is utterly false. IP filtering has been obsolete for quite some time because it is concerned with where an anonymous user is located, instead of who the user is.

Seamless Access seeks to improve remote access scenarios and create a better user experience because IP authentication is very counterintuitive to the current user research experience. It forces researchers to begin their research from, or at some point to navigate through, the university portal to locate the proxy-prefixed URL necessary for remote access. This simply is not how researchers conduct their research. Current obstacles to access include forcing the user to perform numerous clicks to access content behind a paywall, and users typically have credentials scattered over a multitude of platforms that are difficult to manage. If an institution provides numerous solutions for content access (e.g. VPN, EZproxy, Shibboleth, etc.) then users can become overwhelmed with complicated instructions regarding which protocol to select, and how to implement these on their local devices. By providing such complicated procedures for navigating beyond a paywall, libraries may be inadvertently pushing fully entitled end users to turn to alternative resources such as SciHub or ResearchGate to obtain easily accessible content.

IP filtering has also proven to be time-consuming and expensive to manage. The connection details of every provider of licensed content must be registered in the proxy server's control file and maintained over time. If the subscribing institution's IP address range changes, these changes must be coordinated with potentially hundreds of providers. Lastly, IP access 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.

USER EXPERIENCE

Because the overall user authentication experience is currently inconsistent, confusing, and replete with jargon, Seamless Access is implementing a standard for federated authentication based on a single sign on through the user's home institution. Regardless of where the end user begins their research they will encounter consistent imagery, language, and login placement, along with a standardized identity provider discovery flow.

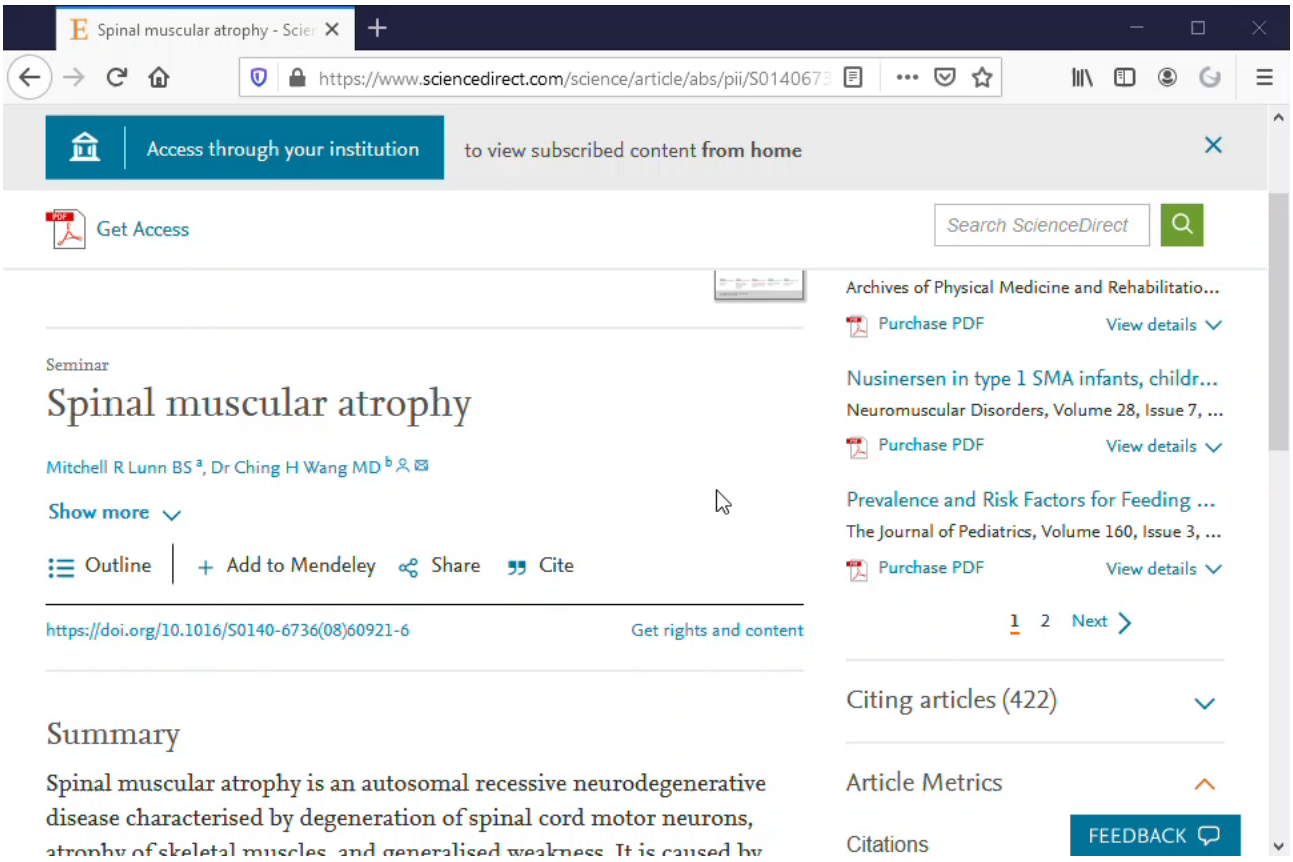

Figure 1. Consistent imagery, language, and placement.

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

ENSURING USER PRIVACY

Privacy is all about balancing security and accessibility. While it's possible to configure access that is entirely anonymous or entirely transparent (to use examples from each end of the scale), the vast majority of users and organizations opt for various shades of grey. In other words, a user may choose to give up some privacy in return for something valuable such as a level of personalization that makes use of a resource more efficient and more engaging. How this plays out in the library world is that each library makes decisions based on institutional and library policies and can also depend on the nature of the resources.

For example, libraries may license resources where access should always be anonymous due to the sensitive nature of the material. But most users value some level of personalization, even if that's simply to retain topics of interest so that they don't have to be rediscovered every time a platform is used. There are also some valid reasons why personal data needs to be shared for some resources, such as when a patron is doing online training and needs to have their learning personally accredited to

them. Therefore, the ideal access solution is one that gives libraries control over what shade of grey is appropriate for their organization and their various types of resources.

Federated Authentication puts libraries in control of privacy. The institution decides what user data (or attributes), if any, are shared with a vendor, and attribute release only occurs after a user is authenticated which gives libraries important control over access, costs, and risks.

DATA PRIVACY & ATTRIBUTES

In federated identity management, attributes is the term used to describe data about an authenticated user and attribute release is the process by which that data is shared as part of the authentication process by an Identity Provider (IdP), such as an institution, with a Service Provider (SP), such as a publisher. The format an attribute takes depends on the underlying technology. For example, Security Assertion Markup Language (SAML) is the technology that underpins Shibboleth and OpenAthens, but there are other technologies that support federated authentication such as OpenID Connect, which is used by consumer-focused services like Facebook and Google.

It's important to note that personally-identifiable attributes are not required as part of federated authentication. An identity provider can simply assert that a user is an authorized member of their organization and do nothing more. In this case, the identity provider would provide an anonymous assertion identifier that would be associated by the service provider; e.g.: *d71a3a8e9fcc45c9e9d248ef7049393fc8f04e5f75.* Since this identifier is uniquely generated for each authentication and contains no personally-identifiable information, it ensures that user privacy is preserved.

Here are some examples of the types of attributes that can be passed as a result of a successful user authentication:

Affiliation	Define the user's association with their home institution e.g. faculty
Entitlements	Confirm the user's rights to access a resource e.g. URL for licensing contract
Pseudonymous	Unique ID for every person & SP Real identity unknown (pseudonymous) Personalization possible
Personal	Name, email address

Figure 2. Attribute types.

- **Affiliation attributes** define the organizational association between the user and their home institution, by means of employment, membership, enrollment in an educational program, etc.
- **Entitlement attributes** confirm the user's right to access a given resource based on criteria previously agreed with the service provider.
- A **pseudonymous identifier** is unique to each person and for each service provider, so it masks their true identity but it does enable that user to be identified by the same service provider the next time that they visit. However, it's important to note that this identifier can't be used to build a pattern of usage across service providers. It can also be used to personalize a user's experience.
- There are also **personally-identifiable attributes**, such as name and email address.

Attributes are important because they give both sides of the authentication transaction greater control. This control can be valuable in a variety of different ways. For example:

- **Access control**: a library can choose to make a resource available only to users who are full-time staff and students, preventing, say, alumni or contractors from access.
- **Cost control**: a library can limit resource access to users with a certain role or from a certain department.
- **Risk control**: pseudonymous IDs allow users to benefit from personalization without exposing them to the risks and inconvenience of separately registering yet another username and password. The service provider can recognize a returning pseudonymous ID and personalize that user's experience accordingly without receiving any personally identifiable data, without needing to store their email address, and without asking for a password.

Attribute release only occurs after a user is authenticated - a service provider can't pull attributes. They only receive what the identity provider chooses to send. It is configured by the identity provider for each category of service provider since different levels of privacy may be appropriate in different situations. Library resource access is only one of a number of valuable use cases for federated

authentication. For example, research collaborations involving researchers across different institutions would typically share some personal data, such as a name and email address. Also, institutional workflows that require users to confirm their institutional affiliation with third parties may involve scenarios where it is appropriate to share a much broader range of user data, such as authorizing the use of institutional funds for open access publishing fees. Because the identity provider is in control, any special needs for attributes need to be agreed upon in advance so that attribute release can be configured appropriately.

But libraries face a challenge when it comes to configuring access. To avoid organizations having to manually configure exactly which attributes to send to each service provider, configuration is managed through entity categories. An entity category is a metadata tag used to group entities like service providers or libraries so that a standard set of attributes can be built and applied at the group level rather than the individual entity level. The most well-known entity category in use today is the REFEDS Research & Scholarship (or R&S) entity category[2]. REFEDS is the Research and Education FEDerations group, which represents the global research and education identity federations. This entity category only applies to service providers that are "operated for the purpose of supporting research and scholarship interaction, collaboration or management" It cannot be used for access to licensed online resources. This means that there are currently no standards for how organizations should release attributes for the many use cases that fall outside of the R&S entity category - such as library access to licensed resources.

SEAMLESS ACCESS WORKING GROUPS

To address this problem, Seamless Access is developing standards for attribute release in the context of library resources. Its goals are to:

- Set broadly understood expectations about user privacy for Federated Authentication for library resources
- Standardize the set of attributes released in typical resource-access use cases
- Simplify configuration for institutions (and avoid mistakes)

To this end Seamless Access created two cross-industry working groups in order to develop best practice recommendations. The first is the Attribute Release Working Group, which has more than twenty members from across industry stakeholders including service providers, libraries, federations, and consultants. Based on the recommendations of the Attribute Release working group, Seamless Access proposed new entity categories and associated attribute bundles that would create attribute release standards for access to library resources:

- The **Anonymous Authorization** category would be used by a service provider who needs to filter access based on a user's affiliation and/or entitlements.
- The **Pseudonymous Authorization** category would be used by a service provider who needs to personalize their service, and would also allow for additional entitlement and affiliation data that could provide more control over access, such as a user's role.

[2] https://refeds.org/category/research-and-scholarship

These categories were recently approved by REFEDS, the Research and Education Federation community organization, who will become the custodians for these entity categories and NISO will endorse them. Also, note that libraries and their vendors can still agree to release additional attributes via bilateral agreements, but it should be a conversation rather than an assumption.

The second working group is the Contract Language Working Group. Its task is to develop contract language templates for library use based on these proposed entity categories, which will give libraries a mechanism to ensure attribute release compliance. This working group recently started their work.

PRIVACY POLICIES

To reinforce its commitment to privacy, Seamless Access is requiring that service providers follow the GÉANT Data Protection Code of Conduct[3] for use of the service. This document provides specific guidance to service providers on how they should handle personal data in the context of federated authentication and covers four fundamental principles: purpose limitation, data minimization, deviating purposes, and data retention. This means that service providers should only use attributes necessary for access, use as little data as possible wherever possible, only use this data to provide access, and delete or anonymize this data when it's no longer needed. It also aligns very closely with the American Library Association's library privacy guidelines found in its Code of Ethics[4].

HOW DO LIBRARIES PARTICIPATE?

As vendors and publishers across the industry increasingly adopt federated access and SeamlessAccess, libraries can participate in a number of ways. Firstly, plan for the transition and find out if federated access is already supported at the institution. Consider what privacy policies should be adopted in relation to attribute release (the process by which user data can be shared with Service Providers), which will likely be guided by policy at both the institutional and library level. If they haven't already done so, librarians should get to know their IT department and be prepared to help them understand the library's needs.

Vended solutions that support library federated access should be evaluated. The underlying technologies are complex and it may not be something the institution's IT department wants to add to their list of responsibilities. And lastly, librarians should get involved if they are interested. A monthly newsletter can be subscribed to on the Seamless Access website, and librarians can consider joining one or more of Seamless Access working groups that are shaping policy and best practices in this area.

[3] https://wiki.geant.org/display/eduGAIN/Data+Protection+Code+of+Conduct+Cookbook
[4] http://www.ala.org/united/sites/ala.org.united/files/content/trustees/orgtools/policies/ALA-code-of-ethics.pdf