Appendix 2a: Technical Recommendation for Authentication and Authorization within Humanities Commons

Introduction
This report provides technical guidance for future stages of development for Humanities Commons, specifically regarding issues of identity management and cross-organizational collaboration. One of the goals of Humanities Commons is the development of a federated network of sites to support and foster online communication; underlying this development must be an adherence to best practices for federated identity management, both to provide a safe and secure space for participants to engage with one another in scholarly pursuits and to ensure the technical flexibility needed for Humanities Commons to pivot or scale as appropriate with future endeavors.

Two potential technical models exist for the future Humanities Commons overall: the hub-and-spoke model and the federation model. Already discussed in the Humanities Commons grant narrative, these two content and interaction models can be summarized thusly:

- **Hub-and-spoke** requires a single hub—a single Commons instance—for the creation and publishing of content through “spokes” that are secondary sites within the overarching instance (or may be standalone instances elsewhere, in an extension of the model).

- **Federation** requires no single hub, but instead relies entirely on technical communication standards and protocols to create the network; each organization would be responsible for the technical installation, maintenance, and management of its own Commons site.

The selection of the future technical architecture—either a pure hub-and-spoke model, pure federation model, or a combination of the two—should be informed by a thorough understanding of each organization’s existing technology and resources. A brief survey of ten potential partner organizations has already shown a wide variety of technology, approaches to data management, and available resources; this survey has led to the recommendation that follows in this report. Specifically, identity management for the Humanities Commons should be aggregated from canonical sources and maintained in a centralized location, abstracted from the content and interaction model (be it a hub-and-spoke model, federation model, or other).

Understanding Authentication and Authorization
Any discussion of identity management needs and solutions should begin with an understanding of the concepts of authentication and authorization, as the differences are quite pronounced yet the terms are often conflated.
• Authentication, also referred to in shorthand as “authn,” verifies you are who you claim to be, through shared secrets such as a password, authentication token, or even biometric data: you are user janedoe@gmail.com logging in to your Google account, or user jdoe123 logging in to your university’s UNIX account.

• Authorization on the other hand—also referred to in shorthand as “authz”–determines the level of access you have for a given system: janedoe@gmail.com is allowed to browse her mail, user jdoe123 is not allowed to read or write in the /var directory.

Regardless of the technological model used for content and interaction, the Humanities Commons will require basic authentication and authorization for participation. For the consumption of content that is not fully open access, potential users will be authenticated and their access to resources authorized. This process is especially true for content creation; only authenticated and authorized users will be able to create or include content within the Humanities Commons framework. Authorization to read or create content within specific resources will also be controlled; for example, authorized users belonging to a specific organization or group or committee within an organization, would be authorized to read or create content constrained only to that group.

The information that follows in the sections regarding the technologies and processes used by member organizations—specifically the heterogeneous mixture of technologies and processes—supports the initial assertion that identity management for the Humanities Commons should be aggregated from canonical sources and maintained in a centralized location, abstracted from the content and interaction model.

Technologies and Processes Used by Member Organizations
Each participating organization in Humanities Commons will maintain its own membership records; the internal definition and ongoing administration of membership records will remain within the purview of each organization. However, in any technological model for Humanities Commons, a basic set of information will be necessary to first authenticate and then authorize users for specific resources.

In order to determine the feasibility of any technical solution, ten organizations provided detailed information about the technology and processes currently used to support their membership databases. This information is summarized in the sections that follow.

Database Software
As expected, the list of databases used by the member organizations varied from a basic spreadsheet in Microsoft Excel to the use of a relational database management system (RDBMS) with custom data schema and integrations. Additionally, 40% of organizations surveyed used an
off-the-shelf commercial customer relationship management program (CRM) specifically for organizational membership management.

- Commercial CRM (netForum, ClearVantage) – 40%
- RDBMS (Oracle, MySQL, MS-SQL) – 30%
- Microsoft Office Suite (Excel, Access) – 20%

The wide variety of database technology used by the organizations surveyed is likely indicative of the range of technology used by organizations overall.

**Membership Database Update Frequency**

The frequency and process by which organizations update their membership database also plays a role in the proposed technical solution for authentication and authorization within *Humanities Commons*. We cannot assume that all membership databases are updated in real-time when a user purchases or renews a membership, or even that all organizations use an online process for membership purpose and renewal. In fact, 20% of the organizations surveyed used an entirely internal, manual, and off-line process for updating their membership rolls; these organizations gather new and renewal membership requests in multiple forms, process the requests manually, and update the database manually on either a daily or weekly basis. The remaining 80% of the organizations have an online process for new and renewal membership requests, which are updated in real-time after a successful payment transaction, with an additional ability for internal employees to update the membership database directly in the case of non-electronic requests (such as a physical check sent through the mail).

**Minimal Required Data**

Because of the need for a basic set of information to authenticate and authorize users for specific resources within *Humanities Commons*, the baseline must be established from the existing set of minimal required data for each organization so that participation does not cause undue burden on the individual organizations through a change in process or technical requirements. A normal set of required data for account-based resource access is:

- email address (or username; usernames are often email addresses)
- password
- first and last name
- affiliation

Every member organization currently requires at least a first name, last name, and email address for all of their users. Members within an organization would first be affiliated with that organization in any future membership system; therefore the affiliation requirement is met. Some organizations (40%) also require some value for an affiliation such as college, university, other organization, or an affiliation of “independent” or “unaffiliated,” which could also be used.
However, as not all organizations have an online presence that requires a password for accessing online materials, therefore not all organizations require a password. The lack of fully-transferrable membership information is certainly not insurmountable, but as every Humanities Commons user must have a password, some bridging of this gap will be part of any technological proposal for identity management.

**Ability to Record Multiple Email Addresses**
Responses were split evenly between organizations who allow multiple email addresses to be associated with each user account. For those organizations that allow multiple email addresses per account, the user must specify a primary account. Regardless of the authentication and authorization solution used for Humanities Commons, there must be an early decision whether or not users can be authorized by one or more email address.

**Optional Required Fields**
Nine of the ten organizations surveyed have a wide range of optional fields that complete a member’s profile in their organizational database. From mailing address to academic position to other demographic information such as gender and ethnicity, organizations capture a range of personal and professional information for internal use. Although some of this information could be leveraged to authorize users to specific areas in Humanities Commons (e.g. Groups for members at the same institution), it would not be appropriate to base any technological requirements on optional fields within membership records.

**Committees and Groups within Organizations**
Half of the ten organizations surveyed have organizational committees (including a Board of Directors) on which users may serve while their membership is active. It is reasonable to assume that organizations who make use of committees in this way will want to carry over this type of structure in Humanities Commons, to authorize users to specific areas used by these groups. Although it would not be appropriate to base technological requirements on optional fields within membership records, it *would* be appropriate to require the authorization and authentication solution for Humanities Commons to employ group-based authorization within the sites or sub-sites used by the organization.

**Size of Membership Database**
To have a better understanding of the number of users potentially participating in the early version of Humanities Commons, each organization reported the number of active and total (active plus inactive) membership records. The active number ranged from 1800 to over 30,000, and the total number from 4000 to over 50,000. A round number of 100,000 total active members from the ten organizations initially surveyed provides a good sense of scale for the next phase of the project.
A follow-up question asked the organizations how member records moved from an active to an inactive state (which could mean anything from an expired membership to a deceased member). For half of the organizations, a specific flag is set on the record in the database indicates the status. The remaining organizations look at a field such as membership year (a membership year of 2013 would mean the user is valid for any date in 2013) or renewal date (a renewal date in the past indicates an inactive record).

**Existing Identity Management Connections**
The final question in the organizational survey asked if members had the ability to login to the organization’s web site or other online properties through existing single sign-on (SSO) or open authentication protocols. As expected, the answer was overwhelmingly “no”, but two organizations indicated an internal identity management program in place such that authenticated members were able to use SSO to login seamlessly to other web properties maintained by the organization.

**Possible Solutions for Authentication and Authorization**
The central question is how to allow organizations to control their membership records, yet allow those membership records to be used without unreasonable delay to authenticate and authorize users to access resources housed within a hub-and-spoke, federated, or blended site model for *Humanities Commons*. The sections below summarize some common technological approaches to this issue, a summary of issues that exist for each approach, and an overall recommendation for use.

The “Organizational Involvement” and “*Humanities Commons* Involvement” descriptions for each approach already assume that organizations will create the content of sites or sub-sites within the *Humanities Commons*, and that the *Humanities Commons* will provide the core platform and mechanism for registration. The descriptions thus encompass additional required involvement.

- **Users create ad-hoc *Humanities Commons* accounts and select own organizational relationships.**
  - **Organizational Involvement**: None.
  - **Humanities Commons Involvement**: None.
  - **Issues**: Creating accounts entirely on the honor system means no real authentication; anyone could just as easily say they are a member of the MLA as they are a member of the AHA. While this is certainly a possible real-world example, without true authentication the ability to authorize to member-specific resources becomes impossible.
  - **Recommendation Status**: Not recommended.
Users create ad-hoc Humanities Commons accounts, select own organizational relationships, and wait for organizational approval.

- **Organizational Involvement**: Requires ongoing review of registration queue and manual lookups from own database for approval.
- **Humanities Commons Involvement**: Requires creation of approval mechanism.
- **Issues**: Each organization would need to allocate resources to this approval queue; without any automated matching to existing member records. Users would have an unspecified waiting period until their registration was manually approved. In this process, there is no concept of the expiration of membership, so users would be authorized for resources in perpetuity.
- **Recommendation Status**: Not recommended.

Users create ad-hoc Humanities Commons accounts based on authorization codes provided by organization.

- **Organizational Involvement**: Requires creation and management of authorization codes.
- **Humanities Commons Involvement**: Requires creation of authorization code management for organizations.
- **Issues**: Each organization would need to allocate resources to the creation of an authorization code scheme (e.g. Bulk assignment and expiration, individual assignment and expiration), and the delivery of these codes to users. Users would be able to login immediately and access the resources authorized for their accounts. Unlike the previous approach, this approach does have the concept of authorization expiration. However, this approach requires the organizations to take on resource management work outside of their core membership systems, and Humanities Commons must build more than a simple approval mechanism.
- **Recommendation Status**: Not recommended. While the process would meet the most basic requirements for authentication and authorization, this approach is not recommended due to the amount of upfront and ongoing work on both sides that would only result in a loose coupling of systems still requiring a lot of human involvement.

Individual APIs to member databases to validate ad-hoc Humanities Commons account creation.

- **Organizational Involvement**: Internal or third-party technical resources to manage the process of creating and maintaining a secure, authenticated, authorized API endpoint for use by Humanities Commons developers to integrate into the user account creation process.
• **Humanities Commons Involvement:** Resources devoted to participating in the planning and development of API endpoints by organizations, as well as development of per-organization API integration for validation during user account creation.

• **Issues:** This option would only be applicable to those organizations that have internet-accessible membership databases (20% of the surveyed organizations manage their membership database via office software on a desktop machine). The technical resources and time devoted by both the organization and Humanities Commons staff would be front-loaded in the project and not insignificant, such that delays would cause a ripple effect for overall project development milestones. Finally, using an API at the account validation stage does not address membership expirations, changes in user information, or changes in group assignments; all of those changes would still need to be managed by the user (now in multiple places) or addressed by the organization in yet another technical project to create an administrative interface.

• **Recommendation Status:** Not recommended.

➤ **Individual APIs to member databases to validate logins to Humanities Commons in real-time.**

• **Organizational Involvement:** Internal or third-party technical resources to manage the process of creating and maintaining a secure, authenticated, authorized API endpoint for use by Humanities Commons developers to integrate into the user account creation and login process.

• **Humanities Commons Involvement:** Resources devoted to participating in the planning and development of API endpoints by organizations, as well as development of per-organization API integration for validation during user account creation and at each login attempt.

• **Issues:** This option would only be applicable to those organizations that have internet-accessible membership databases (20% of the surveyed organizations manage their membership database via office software on a desktop machine). The technical resources and time devoted by both the organization and Humanities Commons staff would be front-loaded in the project and not insignificant, such that delays would cause a ripple effect for overall project development milestones. The addition of API-based validation of logins would mitigate the issues of membership expirations, but changes in user information or changes in group assignments would still need to be managed by the user (now in multiple places) or addressed by the organization in yet another technical project to create an administrative interface.
•  **Recommendation Status:** Not recommended.

- **Member organizations implement a publicly accessible identity management provider platform used to validate logins to *Humanities Commons* in real-time.**
  - **Organizational Involvement:** Internal or third-party technical resources to manage the process of creating and maintaining a secure identity management system that can extend single sign-on capabilities to *Humanities Commons*. Each organization would also be responsible for ensuring information in their member databases was populating the identity management system as well.
  - **Humanities Commons Involvement:** Define the identifying information passed in the single sign-on process (for implementation by each organization), and develop the single sign-on mechanism within *Humanities Commons*.
  - **Issues:** The initial and ongoing technical resources would likely be cost-prohibitive for many organizations, and would surely lock out any organization that doesn’t already have an internet-accessible membership database. Although work by *Humanities Commons* staff would be less that with some of the other options, the technical resources and time devoted by both the organization and *Humanities Commons* staff would be front-loaded in the project and not insignificant, such that delays would cause a ripple effect for overall project development milestones. The use of single sign-on for validation of logins and determination of group-based access to resources would mitigate the issues of membership expirations as well as changes in group assignments.
  - **Recommendation Status:** Not recommended.

- ***Humanities Commons* run an aggregated identity management provider platform used to validate logins to *Humanities Commons* in real-time.**
  - **Organizational Involvement:** Minimal internal resources, needed only to export existing data to provide to *Humanities Commons* staff in a standard format on a consistent interval.
  - **Humanities Commons Involvement:** Internal or third-party technical resources to create an import mechanism, build an import tool or manually provide this service for organizations, ensure the identity management platform runs smoothly on an ongoing basis, and develop the single sign-on mechanism within *Humanities Commons*.
  - **Issues:** The upfront and ongoing burden in terms of technical and managerial resources will fall squarely on *Humanities Commons*. However, taking on this burden would eliminate the need for each participating organization to devote time and resources to achieve the same goal (if *Humanities Commons* pays $n in
software licensing and development costs to produce an aggregated identity management provider platform, it eliminates the need for each organization to pay a similar $n cost). As with the previous option, above, the use of single sign-on for validation of logins and determination of group-based access to resources would mitigate the issues of membership expirations as well as changes in group assignments (both described in previous options). The tradeoff in this option is the simple (albeit manual) process of organizations importing raw membership data on a consistent basis, in exchange for significantly reduced up front and ongoing costs for organizations.

- **Recommendation Status:** Recommended.

**Additional Desired Functionality within Humanities Commons**

During the November meeting of potential Humanities Comments partner organizations, additional information surfaced regarding desired core features of the Humanities Commons platform. The sections below address this functionality in a minimal way; at least two of these requirements are themselves detailed and lengthy developmental projects.

- Organizations need to provide access to a specific subset of content within Humanities Commons sites and sub-sites by users who are not officially members of their organizations.
  - This functionality will require an agreed-upon standard for indicating a user’s status in each organization’s canonical database (e.g. member or non-member), as well as generic user type in the Humanities Commons that matches that user status. The content maintainers within the Humanities Commons sites and sub-sites would then be responsible for enabling access to this user type to the resources they are creating.
  - **Example:** Jane Doe is a member of the AHA and is co-chairing a panel with MLA member John Smith at the MLA annual convention. Both Jane and John should have access to MLA resources specific to the annual convention, as well as the ability to contribute resources specific to their topic area. John would have access to do this as an MLA member in good standing; Jane would need to be a Humanities Commons user with access to a subset of MLA member resources – her only paid connection to the MLA would be via her conference registration, rather than her association membership. This example is but one of the different relationships that will exist between Humanities Commons users, partner organizations, and the content in sites and sub-sites; these relationships are not as clear as “membership equals access”, but will require organizations to indicate full or partial access to Humanities Commons resources. The underlying framework
should support a type of “non-member” user both for login and for content restriction.

- **Interaction with Federated Identity Management System:** Within a system of identities aggregated from canonical sources, the responsibility is on the home organization to maintain accurate records and push the required data to the identity management layer of *Humanities Commons*.

- **Robust user profile creation within *Humanities Commons***
  - This functionality will require a definition of “robust” before technical planning and development can begin, but before that an agreement should be made regarding the “home” of the data within these robust profiles, and the interaction between participating organizations’ membership databases and these extended *Humanities Commons* profiles.
  - **Example:** James Smith is a member of the ACLA and has a *Humanities Commons* account. The ACLA membership database holds only basic contact information such as name, address, and institutional affiliation. The *Humanities Commons* profile may allow James to edit this contact information and provide only a subset (or none) for public viewing, but more importantly may allow James to build an annotated online CV including links to external resources and populated by importing external resources. This basic description of an extended profile already raises issues about the set of information seeded by membership data, and whether or not data updated in the *Humanities Commons* profile should update the memberships database if the fields are present in that database. The example used here is of an organization that uses an offline, manually-updated spreadsheet for member information; the question is the same no matter if the organization uses a networked relational database for member information, or if the organization uses commercial off-the-shelf software for member information – what changes are the technical requirements necessary to support the answer.
  - **Additional Information:** At a former private for-profit company, the author of this report designed, developed, and implemented a robust profile product specifically for academics. This product took three months to design, develop, and release, including a testing period, with approximately 3 FTE devoted to the project. The design phase included significant user experience testing, but also extensive research into the underlying need for an extended profile, how students and scholars would use the profile. What we found during user research was that students and scholars wanted a system that would allow them to create a robust profile much like LinkedIn, but with required content and entry formatting that matched the nuances of an academic CV.
This product was provided to users as a free add-on to an already for-pay dossier management service, and the company found that users could create a robust profile from a blank slate using this system in approximately 5 minutes. If building a profile took more time than that, users did not care to finish it, and once initially created did not spend a lot of time grooming it (instead only updating it as new jobs and new publications warranted). At this particular company, the usefulness of this product for users was in the ability to create a robust online profile for reading and review, and specifically not social interaction. As the product owner, the author of this report argued against the creation of social networking functionality within this private company's online profile offering, because the context of this particular product would be problematic for gaining traction in that uphill climb to produce value—creating a social network within a for-pay service in which the primary product was the confidential transmission of private documents did not seem a strong foundation to begin competing with LinkedIn. However, robust profile building specifically within a product intended to showcase scholarly communication, such as Humanities Commons, would be a strong foundation to providing a useful, robust, information-driven online academic profile. When doing so, I strongly urge the development team to understand the work that has already occurred in this space over the last five years, and leverage the design and interactions of these best-of-breed tools to provide a solid user experience with built-in value to those users, from the outset. Without doing so, what could be a strong feature will likely result in very little uptake and very little return on investment.

- **Interaction with Federated Identity Management System:** Since users will have accounts within Humanities Commons, and identity management only serves to authenticate users, there is no direct interaction with this feature. Any communication of data between canonical sources and the Humanities Commons database would fall outside of the identity management system and should probably use a separate syncing mechanism until such time as a fully-scoped integrated system could be designed and developed.

- The Humanities Commons should provide upload and retrieval mechanisms between it and an institutional repository.
  - This functionality will require additional infrastructure and resources to support the creation and maintenance of digital asset management framework such as the open source DSpace or Fedora Commons software systems. Additionally, for the Humanities Commons to support the baseline technical operations of an institutional repository, decisions regarding access control, metadata schemas, dissemination and preservation copies, and so on should be determined before
technical planning and development begins, as decisions in these areas will determine both the full resource requirements as well as the scope of possible technical solutions. Several technical projects have been in undertaken (and are in various stages of completion) over the last five years to provide a usable and stable ingestion, access, and discovery layer on top of a digital asset management system, and I would urge the development team to understand this technical landscape and the extensible open source tools available before heading down yet another path toward accomplishing a similar goal.

- **Additional Information:** The author of this report designed, developed, and implemented an institutional repository at a research university, and is aware of the technical and institutional requirements, decisions, and resources required to design, develop, and implement such solutions. The institutional repository in question took several years and approximately 6 FTEs to design, develop, and release, but this process was hindered by bureaucracy and quality of resources typical in a large state-funded institution. However, it would be entirely reasonable to assume that designing, developing, and (most importantly) maintaining an institutional repository at the core of *Humanities Commons* will be a significant undertaking in terms of time and resources. Additionally, if the functionality regarding bidirectional institutional repository communication within *Humanities Commons* is to extend to external institutional repositories (such as the institutional repositories of users’ home institutions), then the time and effort involved would exponentially increase. Finally, I recommend the *Humanities Commons* design and development team review all of the public documents available for the technical design and implementation of the Digital Public Library of America, as the underlying technical infrastructure is not entirely dissimilar although the scope of the content may be.

- **Interaction with Federated Identity Management System:** Since users will have accounts within Humanities Commons, and identity management only serves to authenticate users, there is no direct interaction with this feature unless federated identity management is to extend to external institutional repositories to harvest or update information.

**Proposed Technical Solution for *Humanities Commons* Identity Management**

Having described numerous options in the preceding sections, and with additional features and future functionality taken into consideration, these findings support the need for an abstracted authentication and authorization layer, with data aggregated from membership organizations. The figure below provides a high level overview of this relationship:
Having dedicated *Humanities Commons* staff manage a federated identity management system that contains aggregated information from participating organizations will release the financial and technical burden from the organizations—that does, of course, place it on *Humanities Commons*, but it also reduces the overall cost of the project. From the end user’s perspective, a federated identity management system will specifically address the issue of “profile fatigue” and allow them to participate in *Humanities Commons* with the same profile across multiple sub-sites, which in most cases will also be the same login information they already use to access online resources within their organization’s non-*Humanities Commons* web sites.

While there are numerous software packages available for federated identity management and single sign-on implementations, many of these packages are commercial and require expensive licensing (e.g. Microsoft Active Directory Federation Services, Novell Access Manager, and so on). However, two software packages are often found in higher education institutions and related organizations—Shibboleth\(^2\) and Central Authentication Service (CAS)\(^3\)—both of which are open source and free software.

In brief, both Shibboleth and CAS would allow *Humanities Commons* to setup and run identity provider services and service provider services. These services work together in that the identity provider manages the core identity of the user (in this case, aggregated from organizational

\(^{1}\) A rough estimate would be: 0.5FTE dedicated to managing the ongoing process of data imports and updates between organizations and the centralized system, minimal server setup (3-5 hours of systems administrator or developer time) and ongoing bandwidth costs (approximately $70/mo as an Amazon EC2 small instance), plus configuration of the system and the SSO connection to *Humanities Commons* sites (10-15 hours of systems administrator or developer time). If this work was not centralized, one could safely assume that every organization would have to take on a similar burden, with cost and time estimates potentially increased exponentially depending on the existing technical infrastructure and skillsets of the organization.


\(^{3}\) [http://www.jasig.org/cas](http://www.jasig.org/cas)
membership databases), and exposes the information to additional service providers for use in authenticating and authorizing users accessing resources. The following scenario is common in institutions of higher education, and perfectly demonstrates the relationship between federated identity management services:

- A faculty member at an institution wants to access library resources, so signs in at the library site.
- This faculty member finishes the library session and wants to access a LMS to review student work, and finds they are already logged in to the LMS, despite not having explicitly logged in at the LMS.
- The faculty member is in fact logged in to all university resources via single sign-on, which became possible with their first login to any federated resource (in this case, the library).
- The faculty member did not actually log in to the library site, but instead logged in to the federated system through a login form at the library site; the faculty member could have logged in to any other system site first (e.g. the LMS) and been similarly logged in across the network.

Because of the existing installed user base, development community, and tools for third-party integration with systems such as WordPress, I recommend using Shibboleth as the recommended software for federated identity management services. For example, the WordPress plugin “Shibbolize”⁴ is functioning open-source software that already exists to make the connection between the aggregated identities and the underlying WordPress-based system that will run Humanities Commons. Using this plugin in particular, the process of creating users within Humanities Commons from users already managed by the identity management system—and doing so in a way that is invisible to the user—is completely taken care of, including the assignment of roles and group access based on user attributes present in the aggregated profile.

Once a user logs in to a Humanities Commons site, all other sites would be available to them via this single sign-on that has already authenticated and authorized their access. As an added bonus, the use of Shibboleth in particular would allow the Humanities Commons to join the InCommon federation⁵. For an annual fee, Humanities Commons could become a federated identity provider and service provider within another federation, one which includes colleges, universities, and other institutions and organizations related to higher education and research. This membership would open up the possibilities for additional resource access, and additional ways for users to access Humanities Commons without any organizational affiliation.

---

⁴ [http://wordpress.org/plugins/shibbolize/](http://wordpress.org/plugins/shibbolize/)
⁵ [http://www.incommon.org](http://www.incommon.org)
In the first example, with *Humanities Commons* acting as an aggregated identity management provider, pending an agreement with a service provider, the users of *Humanities Commons* could gain authorized access to that service provider’s resources through authentication by *Humanities Commons*. In other words, pending an agreement with Hathi Trust (an InCommon service provider) to offer full resources to authenticated *Humanities Commons* users (if an InCommon identity provider), a member of the ACLA who does not have access to the broad range of Hathi Trust materials through their institution could login to Hathi Trust using single sign-on via *Humanities Commons*.

In the second example, with *Humanities Commons* acting as an InCommon service provider, agreements could be made with individual institutions to ensure their faculty have access to all of *Humanities Commons* regardless of their personal membership in an organization. In other words, faculty at Brown University could authenticate themselves in their own university identity management system and gain access to Humanities Commons resources because they carry with them the same user attributes as a user who has logged in through *Humanities Commons* itself.⁶

**General Technical Implementation for *Humanities Commons* Identity Management**

While any general technical implementation of the solution described above will consist of several pieces, these are neither complicated pieces nor is the solution unique. Instead, the implementation should follow what has become a well-worn path for leveraging scalable cloud services and using standard technology. In fact, the most complicated aspects of this solution will be ongoing human resources and governance issues, and not technological ones.

The core of the technical solution for the proposed *Humanities Commons* network of federated sites is the identity management process, all of which could be handled by the open source Shibboleth Identity Provider and Service Provider software. For ease of maintenance and to eliminate single points of failure, two separate server instances (each replicated in a secondary availability zone) are recommended:

---

⁶ This is a technical solution only and in no way speaks to how this model of access might affect the organization-based orientation, not to mention the governance and financial model, of *Humanities Commons*. 
However, it is not necessary for these instances to be loadbalanced as long as the secondary instances are replicated and always available. Because the resources used by the secondary instances would be minor (as they are “awake” but not actively responding to requests), the costs involved would be less than for the primary instances (using Amazon Web Services as an example for hosted instances).

Note that the always-awake secondary instances could be mitigated by moving MySQL databases from the instances to the Amazon RDS service; these instances could instead be kept dark and simply updated from source control if a failover is required.

For background information on creating a “federation” using Shibboleth Identity Provider and Service Provider software, please see https://wiki.shibboleth.net/confluence/display/SHIB2/BuildAFederation. The general technical process for installing and configuring the Identity Provider software follows this general outline (please see https://wiki.shibboleth.net/confluence/display/SHIB2/IdPInstall for complete information).

1. Create instances (Amazon EC2 or otherwise)
2. Obtain secure certificates (for HTTPS)
3. Install Apache + Tomcat
4. Install Shibboleth Identity Provider software
5. Configure Apache + Tomcat
6. Configure database connections for attribute storage and retrieval
7. Configure general federation metadata
8. Configure basic metadata attributes to provide (expose to service provider)

The storage of individual attributes (for later release to a service provider) should come from the ongoing aggregation of this data from Humanities Commons member databases. The scripts required for data transfer will be custom to each organization, and will likely run the gamut from a batch process script that exports to a file format that can be easily parsed and imported, perhaps to a direct API call to retrieve data. Given the wide variety of technologies and
processes used (and described in previous sections), it is inevitable that resources will need to be devoted to the creation of generalized (but customizable) utility scripts, and a process put in place to perform the aggregation on a scheduled automatic or manual basis.

Once user information has been made available to the Identity Provider, a Service Provider can expose it; this exposure forms the foundation of the authentication process used by the Commons-in-a-Box platform, based on WordPress. The general technical process for installing and configuring the Service Provider software follows this general outline (please see https://wiki.shibboleth.net/confluence/display/SHIB2/NativeSPGettingStarted for complete information).

1. Create instances (Amazon EC2 or otherwise)
2. Obtain secure certificates (for HTTPS)
3. Install Apache
4. Install Shibboleth Service Provider software
5. Configure Apache + Tomcat
6. Point Service Provider to appropriate Identity Provider
7. Test the release and retrieval of attributes from the Identity Provider

Once the Identity Provider is releasing attributes, and the Service Provider is receiving them, then this authentication information can flow to the Commons layer; for more information about this process in general, please see https://wiki.shibboleth.net/confluence/display/SHIB2/FlowsAndConfig. A WordPress plugin
called Shibbolize\(^7\) can be used out of the box, or modified as needed, to leverage the information released by the Identity Provider to the Service Provider to allow a user to authenticate with the Commons resources. Per the Shibbolize documentation, an initial login to the WordPress-based system for a user authenticated by the Identity Provider can also create the initial WordPress account, which removes the extra step for the user. Additionally, the Shibbolize plugin can interpret attributes passed along by the Identity Provider and set user roles (such as administrator, editor, contributor, and so on) based on these attributes. This open source plugin could also easily be extended to add or update information from the Identity Provider concerning organizational membership, committee membership, institutional affiliation, or any other relevant user metadata.

As such, the process of technical installation and configuration of an Identity Provider and Service Provider, and middleware (such as Shibbolize) between the two, is less of an issue than determining the underlying data that the Humanities Commons needs to act upon. In other words, determining the metadata attributes to pass between the systems—and how WordPress-based roles and access responds to these attributes—will be more complicated than the technical construction of the federated identity management system itself, as the latter is a well-worn path.

\(^7\) [http://wordpress.org/plugins/shibbolize/](http://wordpress.org/plugins/shibbolize/)