

Restore Shibboleth Support to dataverse.harvard.edu

Background

On 2015-04-27 (14 days after launch) we converted all Shibboleth ~20 accounts at <https://dataverse.harvard.edu> to builtin accounts. In addition, we sent the affected users an email with the following subject: "Harvard Dataverse Institution Login temporarily disabled."

Plan

- Purse these options
 - Consider reverting to working setup we launched 4.0 with (Apache proxying most of the app)
 - Background
 - We communicated to dvn-auth on 2014-04-02 our intentions to proxy the app behind Apache:
<https://lists.iq.harvard.edu/pipermail/dvn-auth/2014-April/000006.html>
 - On 2014-04-04 we polled the Dataverse community if there are any objections to Apache httpd as a potential requirement for running Dataverse 4.0:
https://groups.google.com/d/msg/dataverse-community/iLs3HXbeU1E/P0_kqg0zAMoJ
 - Status: Progress will be tracked in <https://github.com/IQSS/dataverse/issues/2180> and the main work to be done is to first enumerate [the problems we faced with Apache in front](#) and then come up with and execute [a test plan to confirm having Apache is front is stable](#). Meanwhile we are writing a list of [Questions About Running Glassfish Behind Apache](#) for Mike Croft from Payara in reply to his offer to help at <https://twitter.com/croft/status/598223623401398272>. One modification is to upgrade to Apache 2.4 as agreed during a [2015-06-05 Shibboleth Meeting](#) but unfortunately we are still seeing a 503 error after an [upgrade to Apache 2.4](#).
 - Write our own Service Provider (SP) with OpenSAML
 - Status: We need to decide if we really want to pursue this.
- Try to continue using mod_shib with modifications
 - Try running Apache on a non-standard port (i.e. 9999 or 8181) and getting Shibboleth login working.

- Get <http://shibtest.dataverse.org> set up for testing
- Test against the Identity Provider (IdP) at <http://testshib.org>
- Discussion at <http://shibboleth.net/pipermail/users/2015-April/021153.html>
- Implemented 2015-04-29 in <https://github.com/IQSS/dataverse/issues/2117>
- Status: We are concerned enough that it won't work in production with two load balanced web servers that we backed out of this experiment at <https://github.com/IQSS/dataverse/issues/2117#issuecomment-102027734> . At the very least, the port need to be configurable, rather than hard coded to 8181.
- Keep Shib users on port 8181 after login (hacky, 8181 will be in URL)
- Keep Shib users on new different host with a different hostname such as shib.dataverse.harvard.edu (hacky, we want all user to see the same URL)
- Work with LTS to see if we can get more stickiness even when ports change.
 - Status: We've talked about experimenting with HAProxy for increased stickiness. Initial experiments suggest that HAProxy can maintain session stickiness across ports.
- Try to remove the need for running shib.xhtml on port 8181.
 - If necessary, use headers rather than environment variables even if we'll need to be extra careful to do this securely: <https://wiki.shibboleth.net/confluence/display/SHIB2/NativeSPAttributeAccess>
 - "ShibUseHeaders: Defaults to "Off", this turns on the use of request headers to publish attributes to applications. Use of this option should be avoided. Be sure to review the topic on spoof checking if you enable it." <https://wiki.shibboleth.net/confluence/display/SHIB2/NativeSPApacheConfig>
 - A variation on this is to use the rapache host as a separate single instance of mod_shib and use headers rather than environment variables.
- Can we run mod_shib on a single host?
- Have the app/user take the token from Apache? (I'm not sure what this means.)
- Run production on a single Glassfish node (suboptimal because it makes that node a single point of failure).
- Somehow run mod_shib on a different host (shib.dataverse.harvard.edu or something).
 - "Basically, you are attempting to avoid installing the SP with the application, but that is the entire basis of the SP's design." <http://shibboleth.1660669.n2.nabble.com/Shibboleth-SP-on-different-domain-than-application-t600810.html>
- Clarify what our **constraints** are in developing our Shibboleth implementation: https://docs.google.com/a/harvard.edu/document/d/1BTcYHza_gfpJ0QC3tFwNuM8AuAWm_a7FYEGsvo64joA/edit?usp=sharing

- Investigate non-Apache alternatives for Shibboleth support (not a quick fix)
 - <http://josso.org> suggested by Bob Treacy.
 - <http://projects.spring.io/spring-security-saml/> suggested by Bob Treacy and Gustavo Durand.
 - <https://github.com/onelogin/java-saml> seems new, discovered by Philip Durbin.
 - <https://github.com/picketlink/picketlink> <http://picketlink.org> does SAML, discovered by Philip Durbin.
 - Previously identified alternatives are listed at https://docs.google.com/document/d/1y2axfd_ScmXVICFIV8AuPDdp5xHwTag54pUpVefzs5g/edit?usp=sharing
 - OpenAM
 - OIOSAML
 - Writing our own Service Provider (SP) with OpenSAML
- Reach out to people who run Java webapps behind Apache for Shibboleth support.
 - <http://dataverse.nl>
 - Conversation started at <http://irclog.iq.harvard.edu/dataverse/2015-04-28>
 - They are successfully running their fork of DVN 3.x behind Apache and mod_shib.
 - DSpace users
 - "Before DSpace will work with Shibboleth, you must have the following: 1. An Apache web server with the "mod_shib" module installed." <https://wiki.duraspace.org/display/DSDOC4x/Authentication+Plugins#AuthenticationPlugins-ShibbolethAuthentication>
 - Vireo users
 - "Using Shibboleth with Vireo requires that you have proxy through Apache with Shibboleth installed and configured." <https://github.com/TexasDigitalLibrary/Vireo/wiki/Install>
 - whartung in #rest
 - Chatting at <http://irclog.greptilian.com/rest/2015-04-28>
 - For probably a year he's repeated that we should avoid using Shibboleth the software and implement something in Java using OpenSAML. I think he only has a single Identity Provider (IdP) to concern himself with, however. We'd like to support the ~400 IdPs in InCommon as well as international IdPs through whatever federation.
- Reach out to Dataverse Shibboleth stakeholders
 - <http://lists.iq.harvard.edu/mailman/roster/dvn-auth>
- Reach out to Harvard people who know about Shibboleth
 - <http://iam.harvard.edu>
 - <https://www.abcd.harvard.edu/harvard/groups/abcd-security/>
- Reach out to the Shibboleth community in general
 - <http://shibboleth.net/mailman/listinfo/users>

- Conversation started at <http://shibboleth.net/pipermail/users/2015-April/021153.html> and <http://irclog.perlgeek.de/shibboleth/2015-04-28>