AAI: Experiences with Making a Service Available for Users Worldwide

Jozef Mišutka, Ondřej Košarko, Amir Kamran, Michal Josífko, Pavel Straňák
Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25, 118 00 Prague, Czech Republic
E-mail: {misutka, kosarko, kamran, josifko, stranak}@ufal.mff.cuni.cz

Keywords: AAI, Shibboleth, Service Provider, eduGAIN, SPF

1. Overview

Looking at the current state-of-the-art of the services and frameworks for Authentication and Authorization Infrastructure (AAI) reveals different visions and requirements by the creators and the potential users. We summarize our expectations and requirements together with our experience. The text is divided into blocks containing general guidelines and examples.

2. Brief overview of LINDAT/CLARIN case study

The LINDAT/CLARIN Centre for Language Research Infrastructure in the Czech Republic is part of the CLARIN project. The first pillar of the CLARIN vision[1] generally states that at least academics from Social Sciences and Humanities (SSH) fields from European Union and associated countries should be able to access resources associated with the CLARIN project. This can be seen as the AAI requirement for our services.

In the next sections, we will use the LINDAT/CLARIN digital repository for linguistics[2] as our case study. We rely on the Shibboleth[3] system for AAI that is formed by Identity Providers (IdPs) and Service Providers (SPs), which then form national federations and optionally creating inter-federations. Based on our requirements, we cannot select a set of a few supported IdPs and negotiate specific processing with them: although, this would extremely simplify the solution.

3. Designing your service

There are several crucial design decisions when building a service utilizing an AAI framework.

3.1 Identifying users

In most cases, it is not enough for your service to only know that someone has been authenticated but you need to be able to identify the user. In our repository, it is crucial that we can keep track of users that sign licenses to download specific resources. There are two common attributes available to a SP in Shibboleth that may seem suitable for this purpose[4]: eduPersonPrincipalName (eppn) and eduPersonTargetedID (eptid). There is also a preferred but less used (according to our statistics) attribute called persistent-id but it is usually created from eptid. If the requirement is to be able to identify a user "outside" your service (e.g., do an authorization on a project level where you have multiple SPs) then your only choice is to use eppn. However, according to the specification[5] eppn may be reassigned according to local policies[6]. If a service requires only local unique identification then eptid is acceptable.

It may seem that IdPs are more reluctant to release eppn than eptid because eppn is often not privacy-preserving. Attribute statistics (please, see our set-up) of our SP (2014/09/01) showed that 30 out of 44 IdPs release eppn and 29 release eptid. At least one of eppn and eptid was released by all but four IdPs. This is one of the reasons why LINDAT/CLARIN accepts both eppn and eptid. In addition, LINDAT/CLARIN also accepts emails for identification. Storing the eptid (or better the persistent-id) even with eppn may be useful when you want to e.g., automatically find out whether an account is still functional.

3.2 Requesting attributes

Simply said, you cannot rely only on released attributes in case you have to support a larger number of IdPs and you cannot make specific assumptions about them. LINDAT/CLARIN always requires an email and if the IdP does not release this attribute, we explicitly require it from the user; otherwise the user is not able to proceed with the registration.

The release attribute problem is well known and there are several ongoing activities to improve the situation. The main problem is that even though IdPs and SPs are in the

[1] Mission | CLARIN ERIC, seen on 24/06/2014, http://www.clarin.eu/content/mission

[2] LINDAT/CLARIN Repository Home, seen on 24/06/2014, https://lindat.mff.cuni.cz/repository/xmlui

[3] Shibboleth is an open source federated identity management system. It provides Single Sign-On capabilities and allows sites to make informed authorization decisions. https://shibboleth.net

[4] We are using friendly names of the attributes. However, attributes have different formal names depending on the SAML version used. Correct mapping must be defined in shibboleth SP application (e.g., attribute-map.xml).

[5] Internet2 Middleware - eduPerson Object Class Specification, seen on 25/06/2014, https://www.internet2.edu/media/medialibrary/2013/09/04/internet2-mace-dir-eduperson-201203.html

[6] If the eppn is made of a surname@domain and the surname changes e.g., because of a marriage then eppn can change depending on the IdP implementation. Some IdPs even allow changing the username.

same class (national federation / inter-federation like eduGAIN[7]) they often trust each other only on the technical level (certificate exchange[8]) and there are no strict policies that could persuade or enforce otherwise. At the end of the day, the burden and responsibility lies on IdP administrators that should in reality decide about privacy and home organization policy issues, which is often far beyond their work specification. Moreover, according to our experience it can be quite difficult to get a response from an IdP administrator. National federation could help here but in most cases they do have neither resources nor the authority. It may seem that it should be the inter-federations where most of the problems are visible but it is impractical to contact all the IdPs. Many IdPs understand that without additional trust the Shibboleth project is almost useless and they release common attributes.

Sadly, the technical trust is also the case with eduGAIN. One step into the more applicable direction is the GÉANT Data Protection Code of Conduct (DP-CoC)[9] that requires proclamation from the SP to not misuse released attributes but on the other hand the IdP will send all requested attributes. At the moment, it seems the adoption is rather slow; but with major legal issues covered it may very soon gain momentum. However, the IdP still specifies which attributes to release (if any) so there can be again problems with the choice of the correct attribute set to release again.

LINDAT/CLARIN does not rely on any attribute; however, we have implemented the DP-CoC that has already improved the release attribute by at least two IdPs. In certain instances, where the authenticated users could not access our SP, we have directly contacted the concerned IdPs and solved the problem in two cases. There is one more approach that could theoretically resolve the attribute release problem (and also project wide authorization) and that is the usage of a Virtual Organization (VO)[10] platform (e.g., HEXAA, REMS, OpenConext, Perun, UNITY). However, these approaches have to identify a user (which is mostly done using eppn) and the problems described in the previous section apply. In LINDAT/CLARIN, these platforms could be used as an additional source of attributes but at least with some of the platforms it may be problematic because of the process flow.

3.3 Homeless users

Users without conforming IdPs or users from countries that do not have a national federation (thus, no IdP) should have means to authenticate into your service if necessary. Depending on the requirements, there could be one central project wide "homeless" IdP that a SP can trust or otherwise a local authentication system can be provided. In CLARIN, we have a central "homeless" IdP and in LINDAT/CLARIN we make use of it. In addition to that, we also support local user accounts.

4. Getting external users

In order to get external users using Shibboleth, you have to get appropriate server, install shibboleth SP and create metadata of your SP in conformance with your national federation guidelines[11]. You may also want to support DP-CoC from the beginning, so prepare the metadata accordingly. Once you get into your national federation, you should make the list of possible IdPs available in your service user interface. You can either have a designated page for this purpose or directly show the list on your service main page.[12]

If you plan to host only a few services, you may want to use an umbrella SP with proxying to the services, which would ensure one id in all your services. The next step is getting users outside of your national federation, which can be done through inter-federations. One possibility is to use eduGAIN service and/or CLARIN SPF (Service Provider Federation) [13] if you are eligible. Using eduGAIN often means very small change to your metadata and adding the appropriate metadata feed to your SP shibboleth harvested feeds. Using CLARIN SPF you get more European coverage and very likely better position in attribute releasing. According to our experience, IdPs release more attributes to SPs in the same national federation. SPF basically propagates your SP metadata to national federation in fact bypassing IdP "non-national federation" filtering rules. It is also more feasible that a national federation tackles the release attribute problem.

5. What to really expect in real life

We have encountered a set of problems while operating our SP that we will describe below. To detect problems, we have implemented an advanced log parsing so that we could easily inspect the sessions and identify e.g., when a user selected an IdP but never returned back to SP or the difference between attributes sent to the shibboleth SP application and the attributes propagated to our service.

The first set of problems arose from an incomplete attribute filtering (attribute-map.xml). We have encountered some IdPs that use SAML[14] version 1 and

[7] GÉANT: eduGAIN, seen on 24/06/2014, http://www.geant.net/service/edugain/pages/home.aspx

[8] IdPs use a certificate to verify communications with the SPs. You must install the same certificate on both the IdP and the service instance.

[9] GÉANT Data Protection Code of Conduct (DP-CoC) https://refeds.terena.org/index.php/Data_protection_coc

[10] A Virtual Organization (VO) is a group of individuals that have something in common, for example collaborating on a project.

[11] Consider not using the current hostname in your entityID, in case you plan to change the hostname of your SP. Although, entityID is only a string it may be stored with some IdPs to grant you special behavior.

[12] See LINDAT/CLARIN AAI library for more details, http://hdl.handle.net/11234/6-AAI-LIBRARY

[13] CLARIN SPF, seen on 24/06/2014, http://www.clarin.eu/content/service-provider-federation

[14] Security Assertion Markup Language (SAML) is an XML-based standard for exchanging authentication and

others that use SAML version 2, as both versions are incompatible with each other thus need separate mappings.

Even though some attributes should not be multi-valued, however, your service should be able to handle any multi-valued field e.g., in our experience some IdPs return multiple email addresses for users. There will also be cases where IdPs will simply send you no attributes at all. You should try to identify these and ask the users who tried to log in to contact the IdP or you can do it on their behalf.

There will be IdPs with misconfigured metadata so be prepared that not everything is your fault. For example, scoped attributes require the definition of the scope in the IdP metadata otherwise your SP shibboleth will ignore them. The SP behavior is correct because of security issues e.g., impersonating.

There can be users who will never return back after selecting an IdP because the IdP will not recognize your SP mostly because it does not consume appropriate feed. Your SP should be able to hide them from selection. Because of this, we have developed a QA tool for SPs[15], which processes IdPs from a given list, tries to get to the login screen and return the problematic IdPs. There are around 80 IdPs not working in LINDAT/CLARIN due to eduGAIN errors.

6. Acknowledgements

This work is part of LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

authorization data between IdPs and SPs.

[15] See LINDAT/CLARIN AAI Shibboleth QA for more details, http://hdl.handle.net/11234/6-AAI-SHIB-QA