DASISH

Data Service Infrastructure for the Social Sciences and Humanities (DASISH) brings together all 5 ESFRI research infrastructure initiatives in the SSH area. The goal of DASISH is to determine areas of cross- fertilization and synergy in the infrastructure development and to work on concrete joint activities related to data, such as data access, data sharing, data quality, data archiving and legal and ethical aspects.

Authentication and Authorisation Infrastructures in the SSH*

What AAI do SSH Users Need?

An Authentication and Authorization Infrastructure (AAI) is required whenever access to resources needs to be limited to specific (groups of) users. Imagine that you are a researcher from a university and you would like to access data from a CESSDA center in Sweden and, at the same time, access data from a DARIAH archaeological database in Germany. It would be ideal if you could access all services and data of these, and other SSH centers in Europe, with one and the same user identity. Preferably the one provided to you by your own university.

If your institute is part of a so-called trust federation and uses the appropriate technology, such a single sign-on (SSO) scenario is possible and the identity provided by your home organization can be made to allow you access to services offered by others participating in the trust federation.

The Technology

A well-established technology that makes SSO scenarios possible is based on the SAML-protocol. An AAI infrastructure based on this protocol distinguishes Identity Providers (IdP), Service Providers (SP) and discovery services. An IdP is a type of service that takes care of user authentication, and also provides user attributes such as unique identifiers or an email address. IdPs are typically operated by the user's home organization like Universities. SPs are operated by e.g. publishers or data archives and provide relevant services to users using software that integrates such services in a trust federation. During the login procedure the SP redirects a user first to his home organization for authentication by using a discovery service that allows users to specify their specific home organization.

Such a federated model for collaboration between organizations responsible for user identities and service provisioning is referred to as Federated Identity Management (FIM).

The Identity Federation Landscape

Nearly all SSH centers in Europe (can) participate in a national academic trust federation or national Identity Federation (IDF). Typically, in a trust federation a legal contract or policy agreement holds that specifies the member parties and especially prescribes how user identity information should be securely handled and exchanged.

Depending on the federation policy, users have access to all services in the federation by default. However, IdP and SP operators can always overrule the default access policy.

The eduGAIN Inter-federation

The existence of national IDFs is sufficient for connecting users and services within a single country only. To connect users to services abroad, an IDF should participate in eduGAIN, the IDF inter-federation that connects users and services across federations.

Unfortunately, eduGAIN's effectiveness is currently limited. Because of problems associated with cross-border transmission of personal data, most national IDFs have an opt-in policy, requiring IdP operators to specifically agree to connect to another federation's SP via eduGAIN. Such a policy does not scale well for SPs trying to connect to a broad group of users. Therefore initiatives are underway to promote a default opt-out policy or to otherwise convince the IdP operators to allow cross-federation access

* Deliverable D5.1a of WP 5: Trust and PID services Federation Report. To appear at http://dasish.eu/publications/projectreports

One of these initiatives is the 'data protection code of conduct' (Data Protection CoCo) that specifies behavioral rules for the participating SPs with respect to user attributes handling policy. It is hoped that IdP operators will more easily allow their users to connect and release user attributes to SPs that have complied with the Data Protection CoCo.

The CLARIN SPF

Trying to avoid the scalability problems associated with the eduGAIN opt-in policy, the CLARIN project established a trust federation of CLARIN service providers (SPF) that directly makes contracts with the national IDFs. Thus enabling the users from all those IDFs to connect to all CLARIN SPs. The CLARIN SPF currently connects to 6 national IDF's with 261 IdPs, which is currently more than can be achieved via eduGAIN. Evidently there are also overhead costs related to managing a separate federation. CLARIN also operates its own discovery service and an IdP service for 'homeless' users without a proper academic home organization.

The CLARIN SPF is successful, but if the national IDFs in eduGAIN would change from opt-in to an opt-out policy, CLARIN would also see the eduGAIN model as preferable, since it will minimize administrative efforts.

AAI in DARIAH and CESSDA

DARIAH does not plan for a separate federation; instead it promotes inter-federation via eduGAIN. For those users whose home organization does not connect to eduGAIN or blocks access to SPs in other federations, DARIAH operates an IdP for the homeless that has special facilities to distribute the management effort over the different national DARIAH projects.

CESSDA prefers and is investigating solutions for access to protected data based on services offered by national organizations such as the national IDFs. An additional requirement is the need to operate AAI services by relatively small centers with limited IT expertise.

A Common SSH Approach?

From the analysis DASISH made of the different SSH AAI approaches, it is clear that currently there is no place for a separate SSH trust federation. The main reason is that at the moment both CLARIN and DARIAH are content with their current approaches and CESSDA will likely adopt the eduGAIN strategy, just as DARIAH.

Also, a common trust federation for the SSH, separately from a common approach for the whole research domain, is not so relevant if there are no special specific shared SSH AAI aspects. Currently, useful shared aspects only exist as cost savings by operating common federation facilities as homeless IdPs and discovery services. DASISH facilitated discussions that can lead to sharing such facilities, also beyond the SSH. DASISH therefore recommends continued participation in research domain wide discussions as in FIM4R and the RDA FIM interest group**.

Results and Recommendations

DASISH organized several high-level discussions with its partners from CLARIN, CESSDA, DARIAH and representatives from the national IDFs and eduGAIN. This resulted in a report*** which results can be summarized as:

1. DASISH supports eduGAIN's Data Protection Code of Conduct as a means to connect more IdPs.
2. National IDFs joining eduGAIN should have an opt-out policy with respect to users connecting to SPs in other federations.
3. CLARIN will offer the use of its SPF for SSH centers, provided they will also certify as a CLARIN-T (trusted) center.
4. DARIAH will make its homeless IdP available for all SSH users.
5. General AAI service organizations should be encouraged to offer 'FIM services as a service' such as discovery and homeless IdP services.

Further planned collaboration between CLARIN and DARIAH is the mutual recognition of each other's homeless IdPs. This will limit the overhead for 'homeless' IdP operators and users.

** Please refer to https://indico.cern.ch/event/301888/ and https://rd-alliance.org/groups/federated-identity-management.html
*** Please refer to http://dasish.eu/dasishevents/aaiworkshop/Report_on_the_DASISH_SSH_AAI_strategy_meeting_V3.pdf

DASISH

SEVENTH FRAMEWORK PROGRAMME