EGEE USER FORUM

Contribution ID: **154**   Type: **Poster**

Maintaining Corpus Download Facilities with Changing Technology

Monday 12 April 2010 18:51 (3 minutes)

Introduction of federated login using Shibboleth for a linguistic data archive created a problem by making existing local tools for downloading data-sets stop working. To address this problem the use certificate based authentication in combination with a SLCS service was setup. The application domain demanded that special attention was paid to shielding the user from the complexity of working with certificates and making the SLCS handshake as easy as possible. Parallel access to the archive using both Shibboleth and client certificates required careful configuration.

Detailed analysis

At the Max-Planck Institute for Psycholinguistics (MPI) [1] a large set of linguistic corpora from MPI researchers and also external projects is archived and made on-line accessible. Already some years ago, also within the linguistic community, we saw the need to integrate the existing archiving infrastructures, since this would be the basis for a viable e-Science infrastructure for the linguistic domain. A small EU project of four archiving institutions "DAM-LR"[2] was created that aimed at integrating the archives at different levels, including AAI. A requirement was formulated for federated login for all the archives'users such that for instance SSO for distributed collections would become possible. To realize this, the DAM-LR identity federation was created and the archives Shibbolized their web servers. A consequence of this was that local tools that had been used to copy data-sets from the archives to local storage stopped to function since Shibboleth only addresses access by web browsers. This was an unsatisfactory situation especially since further big EU integration projects in the linguistic domain like CLARIN [3] all plan to use federated login and Shibboleth.

Conclusions and Future Work

Currently we have a working application, and the developed software can also be used to enable other tools to access resources in a similar way. Applying this in a EU wide context like the CLARIN project, imposes the question of the status of the SLCS service. Should it be organized on the basis of national NRENs, on a EU wide basis or as part of a virtual organization platform? This is of course a question that may be of interest for similar integration projects.

Impact

Together with the BiG Grid project [4] and SURFnet [5], a project was setup to test if the use of SLCS obtained certificates [6] could be a solution for this problem. In this context, SURFnet set up a SLCS service, accessible by the members of the SURFfed identity federation. Secondly the MPI's repository apache server was configured with mod_ssl and mod_rewrite to allow client certificate-based authentication in parallel with Shibboleth based authentication. Thirdly, the "IMDI-Browser"a local tool that was originally used to download data sets from archives running the LAT archiving software was modified to perform a handshake with the SLCS to obtain the certificate and use it to download the items of a data set.

[1] MPI archive, http://mpi.nl/resources/data
[2] DAM-LR, http://www.mpi.nl/DAM-LR

[3] CLARIN, http://www.clarin.eu
[4] BiG Grid, http://www.biggrid.nl
[5] SURFnet, http://www.surfnet.nl/
[6] SLCS, SWITCH SLCS service, http://www.switch.ch/grid/slcs/

Keywords

Authentication, Shibboleth, SLCS, archiving, linguistics

URL for further information

http://www.nikhef.nl/pub/projects/grid/gridwiki/index.php/User:Msalle

Authors: Mr BROEDER, Daan (Max-Planck Institute for Psycholinguistics); Mr SALLÉ, Mischa (NIKHEF); VAN VALKENHOEF, Tobias (Max-Planck Institute for Psycholinguistics)

Presenter: Mr VAN UYTVANCK, Dieter (Max Planck Institute for Psycholinguistics)

Session Classification: Poster session

Track Classification: End-user environments, scientific gateways and portal technologies