Deliverable 9.2
Distributed Solution Report

DAM-LR

011841

Distributed Access Management
for
Language Resources

implemented as
Specific Support Action

Contract Number: *011841*

Project Coordinator: Peter Wittenburg

Project Web-Site: www.mpi.nl/dam-lr/

Deliverable: D9.2

Authors: MPI

Responsible: MPI

Date: 1.11.2006

Content

1 Introduction

This is a report about the state of affairs with respect to the implementation of the various technical federation components in DAM-LR and about the federation discussion. We describe first the state of the reference implementation developed at the MPI followed by any differences in the local implementations used by the other DAM-LR partners. The basis for this work are the specifications described in D8.x and D9.1.

The metadata situation is described in D3.x. Here we only want to mention that the repositories from MPI, INL and Lund have a joint metadata domain and that IMDI is used as common platform. Currently, MPI is harvesting all relevant contributions and adding everything to its on-line browsable and searchable catalog. The situation at SOAS is complicated in so far as SOAS had a recent server crash, resulting in a gap in the joint metadata domain.

2 PKI System

In December 2006 SOAS and MPI were able to use official EUGridPMA based PKI certificates. Lund and INL had the process almost completed. For testing purposes “proxy certificates” are used until valid ones are available. For the implementation work in DAM-LR this does not form a problem, however, by the end of the project the proxy certificates have to be replaced by official ones. As described in D9.1 obtaining one is a prerequisite for transparent working of the Shibboleth middle-ware.

3 Unique Identifiers

All partners have agreed to use the Handle System (HS) from CNRI for the purpose of identifying archived resources. Since the initial DAM-LR agreement CNRI has implemented a new license policy requiring new applicants for a handle prefix to pay a small annual sum. This did not change the acceptance of the HS, as all DAM-LR partners accept this change in policy.

The HS system has currently been integrated into the reference implementation developed at the MPI in such a way that:

1. A Handle Server (for handle to URL handle resolving) was setup.
2. New ingested resources are associated with handles.
3. A tool is available that allows archive managers to move or rename resources and that will update the HS database with the new URLs.
4. Handles can be used to retrieve resources from the archive.
5. The archive catalog allows users to obtain the handles for the stored resources.

It was agreed earlier to use the HS to exchange authorization records. This would become a necessity when DAM-LR partners store copies of each others resources. This is something not immediately foreseen as a deliverable of the DAM-LR project, but something that will be one of the following logical steps in a federation. At the technical-meeting in November 2006 a preliminary format for disseminating authorization records was agreed upon.

The state for the other partners is:

- Lund uses the reference implementation as described above.
- INL has obtained a handle prefix, has set-up a handle server and issues handles for resources. These handles are not yet given to the users which does not form a restriction for DAM-LR.
- At SOAS a handle prefix was obtained and a handle server setup. The integration within their own local implementation is not yet finished.

Further we can state that:

- All partners defined a postfix syntax.
- Mirrors will be setup at a later phase in 2007.

4 Authentication

Authentication involves two groups of applications or functions. The first group is web applications that are used to manage and access resources. The second is when accessing a resource directly via its URL.

As described in D9.1 authentication components should be integrated with the Shibboleth middleware so the task of authentication will be performed by the home institute of a user when he accesses resources or applications from other DAM-LR partners. The use of Shibboleth also means that authorization or at least the authorization for required transfer of user attributes becomes mixed with the authentication issue.

As explained in D 9.1 OpenLDAP was chosen as an authentication component for the reference implementation. The OpenLDAP server has been installed and is operational in the test environment of the MPI. An LDAP schema was implemented according to the specifications describing in D9.1. It is clear that because of the sensitive nature of authentication and authorization we only can implement these components in our production environment if we have passed extensive testing. It was recognized that within the DAM-LR federation we would need a unique user identifier space to administer each other. A simple way to achieve this is to use an organizational id as prefix to each users (local) identification separated by a semi-colon.

The synchronization of the MPI's own local user records, currently stored within Active Directory Service, with the LDAP is performed by a MBean application, that regularly copies the records of the MPI's local users to the LDAP. The list of copied attributes and their new names can be easily changed while the synchronization period is a configurable parameter.

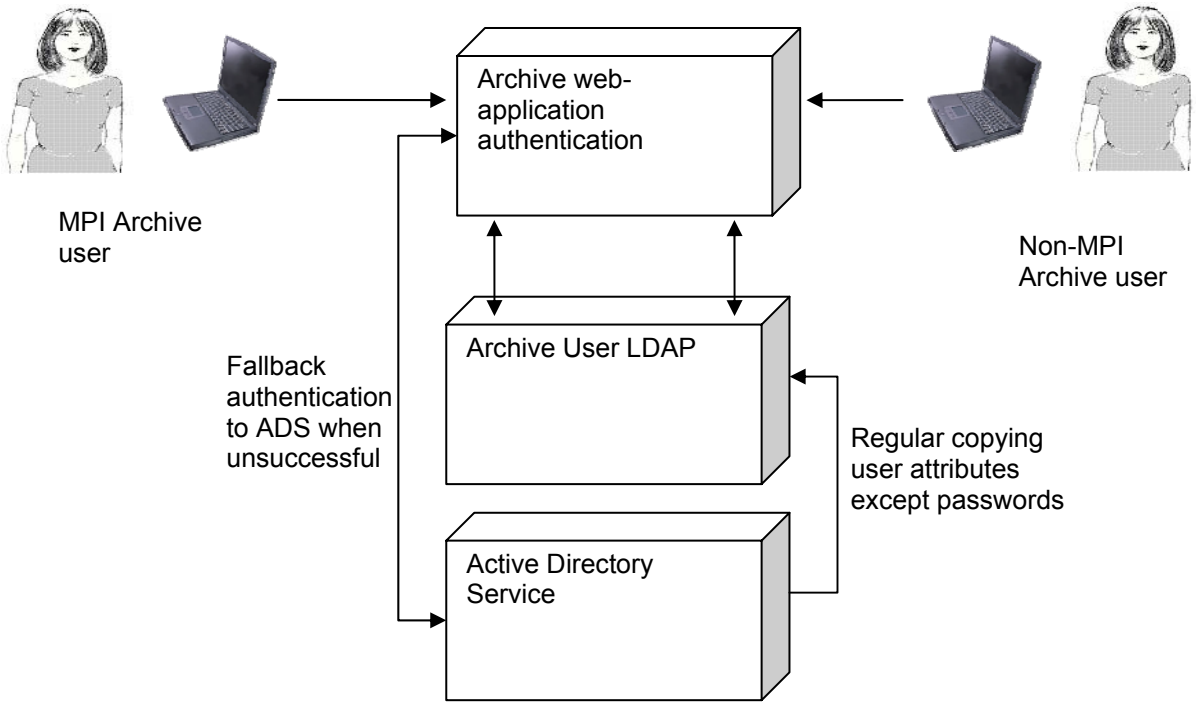
The archive web-applications of the reference implementation will authenticate and obtain user attributes using an API to access the LDAP. The API has been implemented and allows authentication but does not yet offer access to the complete set of user attributes. The current local setup of authentication with MS Active Directory Service for the MPI local users precludes simple copying of passwords to the archive user records LDAP. Special programming logic takes care that for authenticating users without a password present in the LDAP, the authentication is done via Kerberos with the ADS.

When accessing resources directly via their URL, the authentication and authorization mechanisms of the Apache HTTP web-server are used. Shibboleth offers a SP (server provider) component that is integrated with the Apache web-server and that will redirect users to authenticate with their home organization. The redirection mechanism has been shown to work (tested between MPI and INL) but allowing access on the basis of transferred user attributes remains to be implemented until February.

The Shibboleth IP (Identity Provider) component is where authentication requests from the SP are processed. In the MPI setup we need to be able to create the same fall back mechanism for local MPI users as when authenticating for web-application as described above. This is arranged by making the IP use the security mechanisms of a Tomcat Servlet container that is configured to use a JAAS realm. The JAAS realm allows falling back to ADS authentication if authentication with the LDAP fails, just as is the case with the authentication API. Configuring the IP to use the JAAS realm is work in progress.

The state of affairs with respect to the other DAM-LR partners

- Lund will use a copy of the MPI reference setup
- INL has a working IP, but is waiting for a more complete MPI setup to take over solutions.
- work at SOAS is currently behind schedule due to the server problems



5 Authorization

The Shibboleth SP that is connected with the Apache web-server supports the use of authorization records in the same fashion as the normal Apache web-server, which in the case of the reference implementation is storage in the htaccess file(s). However the operation within the DAM-LR federation requires that some extra issues need to be addressed:

- In the authorization records we need to use the federation wide unique user identifiers.
- The IP needs to provide as the user identifier the federation wide unique one. This can be achieved by configuring the IP such that it creates the user identifier attribute by concatenating the appropriate user attributes from the LDAP.
- In the reference implementation authorization records are created by an existing software component: the Access Management System (AMS) that must be able to draw on an existing list of potential users. However non-local users should be added to this list when necessary but only when their credentials are trustworthy. This can be achieved by a function of a new component protected by Shibboleth, where a user can specify his particulars and that is then able to check if these match the information given by the remote Shibboleth IP.

All this is currently work in progress.

6 Federation Aspects

After having discussed the first note about federation issues intensively at various occasions (RI Workshop at LREC, DELAMAN workshop, DAM-LR workshops), a new paper has been created that relates the set of rules needed for a federation with the architectural solution chosen. This will be discussed in the annual report.

7 First Experiences

After one year of hard implementation work much experience has been gained which should be explained. This will be discussed in the annual report.

8 Further Work

Our primary objective is currently to complete the Shibboleth-based configuration:

1. Create or complete the necessary configuration files for exchange of the agreed user attributes

2. Complete and share the Shibboleth's metadata.xml file, identifying all the DAM-LR partners with their certificate keys.

Also the reference model at the MPI will be extended with the Resource Request System (RRS) allowing users from the other DAM-LR partners to post access requests.

The shared metadata domain is to be finalized by having those partners that are able to house an IMDI metadata catalog, harvest each others metadata as the MPI is already doing, i.e., to have full portals also at Lund and INL.

If time permits we could invest some time for the following subjects that could also be subject of a future DAM-LR II project:

1. Investigate and create an efficient way for copying resources or synchronizing whole corpora between archives. Here we should address complex issues such as administrating ownership, access authorization (of which the theoretical planning is part of DAM-LR) and administrating the URIDs for copied resources.
2. Make the shared metadata domain more efficient by making the different metadata catalogs housed by partners that copy the MPI reference model "complementary" rather than "duplicated". This would entail being able to parallelize metadata search queries over different distributed catalogs and being able to configure complementary domains of authority for a metadata catalog, although overlap is not harmful.