PaN-data ODI

Deliverable D3.2

Pilot Deployment of initial AAI service infrastructure

Grant Agreement Number	RI-283556
Project Title	PaN-data Open Data Infrastructure
Title of Deliverable	Pilot deployment of initial AAI service infrastructure
Deliverable Number	D3.2
Lead Beneficiary	PSI
Deliverable Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	01 Oct 2012 (Month 12)
Actual Delivery Date	31 Oct 2012

The PaN-data ODI project is partly funded by the European Commission under the 7th Framework Programme, Information Society Technologies, Research Infrastructures.

Pilot deployment of initial AAI service infrastructure

PaNdata / WP3

Draft

Version 2.0, hjw Oct 28, 2012

Rev Oct 30, 2012

1 Introduction

The European large photon and neutron infrastructure provide their users with research possibilities of highest scientific level. This is reflected by many Nobel prizes awarded in recent years to projects related to research performed at these facilities.

In recent years, there is an increased need for facility services to this community on a transnational scale, beyond the scope of single infrastructures, especially access to remote IT services. A closer look reveals that all these services require a common, pan-European AAI service. From that, it becomes clear from the beginning, that such a service needs two important elements, (a) a solid technical concept and (b) a realization / implementation scheme which is common for all European facilities.

2 The Umbrella concept

2.1 History and general design

Umbrella was originally developed between 2008 and 2011 within WP2 of EuroFEL as a prototype for the web-based access point for large photon / neutron infrastructures [1]. It was further developed in the PaNdata Europe project between 2009 and 2011 and in now being extended in PaNdata ODI and CRISP projects. The Umbrella design for an AAI system for European large photon and neutron facilities users is the consequence of several basic requirements.

2.1.1 Central identity provider and local user offices

All large research infrastructures have their own web-based user (WUO) offices which manage their part of the 30’000+ user of these facilities in Europe by means of own user databases. The idea of Umbrella is to link these local user databases together through one central identity provider. This will enable identification of the user accounts which belong to the same person at different facilities and so support users who wish to bring together their data collected at separate facilities. It will also enable single sign on. Technically this is relatively straightforward, as the requirements of the different user offices are very similar, which leads implicitly to very similar database structures. Umbrella is designed to be flexible, with e.g. identification strength depending on the specific requirement.

There is a wide multi-national distribution of the large facility users. Thus, with a nationally structured identification system many different identity providers will be involved. As within Umbrella this identification is handled mainly by Umbrella and the facilities in cooperation, this is easily organized, despite the fact that the users originate geographically from very many states. However non-technical requirements such as privacy and data protection concerns can make the implementation of technical solutions considerably more complicated. As an example, technically the simplest solution for a user database would be to store

all user information in a single central database. This is the solution chosen for many academic and commercial systems. For the photon / neutron large infrastructures, however, confidentiality between facilities and users plays an important role. In a central system, however, confidentiality protection, especially within the system is always limited. In the Umbrella system only that part of the user information is stored centrally, which is necessary for the identification of a user. All other information is kept at those infrastructures where the user has registered for services. In order to arrive at the necessary functionality, additional synchronization mechanisms (pull only) have to be implemented.

2.1.2 Unique persistent user identification

Increasingly, remote IT actions like remote data access or remote experiment access are becoming important. These services are possible only by the unique identification of a user. In addition, as many of these users are PhD students and postdocs with a high chance to change affiliation after few years, this identification must be persistent, i.e. affiliation independent. Within the Umbrella system this is given by having a single identity provider.

2.1.3 DIY mechanism, data protection and user-initiated actions

The user community of the European large facilities has a highly international character and IT activities in the various states are subject to strongly diverging legal boundary conditions. Data protection and privacy legislation within member states impose far-reaching conditions on the exchange of data about users between facilities. Within a general top-down ‘identity-provider’/ ‘service-provider’ scheme this leads to complex trust agreements between the participating partners. Within Umbrella, attribute exchange is handled in a bottom-up scheme, where the user him/herself triggers the transfer of his own information (user-initiated or so-it-yourself (DIY) actions. For resource-requesting actions, e.g. involving licenses, verification of user-supplied information may be required. This could be implemented easily via additional handshake mechanisms.

2.1.4 Hybrid data structure

The databases of the Umbrella system are defined according to a hybrid local / central format. The idea is to keep most of the information locally at the facility user office databases and centralize only those items which are necessary for a federated operation. Specifically, concerning the user database, all user information is kept locally, as it is now, except for that information which is necessary for the user identification (username, password). The latter information is stored in the central identity provider. Clearly, an additional synchronization mechanism is required for keeping the local databases in sync. This is accomplished by the DIY mechanism described above (section 2.1.3). If a user is e.g. registered at several WUO systems and needs to update his/her personal information, he/she retrieves the information from one of the WUO systems, edits it and distributes it via a specific Umbrella tool to the other WUOs.

As a secondary product of this structure there is inherently a very loose connection between the WUOs and the Umbrella system and that each WUO can decide on its own when to join the common Umbrella system. In this way the adoption of Umbrella can be incremental, with no need for a “big-bang” introduction of a new Database.

2.1.5 Single sign on

Besides providing a unique, persistent user ID, the single-sign-on capability is the most important capability of Umbrella, so that users are no longer forced to login to a multitude of different services, all with different identifications. Based upon the Shibboleth technique (see 2.2 below) users log in once and have access to all services offered within the Open Data Infrastructure across the participating facilities.

2.1.6 Slim concept, coexistence with local WUOs

By its concept, Umbrella is designed not as a replacement of the existing WUO structure but rather as an extension. In view of the very large number of users the local WUOs do an excellent job and there is no need for a principal change. The introduction of Umbrella will require some additional resources but these will be compensated by synergies as several actions can now be performed in common.

2.2 Technical realization

The Umbrella authentication system is built upon a user directory for the storage of the accounts and a SAML2 ecosystem, which enables Single Sign On. SAML2 is an OASIS standard [2]] and has wide support in academia (Switzerland: SWITCHaai, Germany: DFN-AAI, Denmark: DK-AAI) and industry (e.g. AOL, EMC, HP, IBM, Microsoft, Nokia, Oracle, Red Hat, Boeing). Finally, it has been decided to use Shibboleth2 [3] by the Internet2 Middleware Initiative as the implementation of the SAML2 specification. As outlined in section 2.1.1, a specialty of Umbrella is, that it has only one (1) identity provider. This allows a unique identification of all users of the participating infrastructures.

A detailed technical description of the Umbrella system is provided in EuroFEL deliverable 2.5 [4].

2.3 Umbrella development in PaNdata and CRISP Projects

Because it underpins many other services, identity management has become an important topic not only of PaNdata, but also several other EU projects (e.g. NMI3, CALIPSO, CRISP), where the strongest topical overlap is between PaNdata and CRISP. In order to make best use of the available resources, and to avoid any double activities the PaNdata and CRISP responsibles have joined synergies from the beginning in order to coordinate their activities. Amongst others, this has resulted in three harmonization workshops (Zürich (June 2011 [5]) and 2012 [6]) and Hamburg (December 2011 [7]). These workshops were

attended by over 20 participants, representing the European large photon / neutron infrastructures, and established a common understanding and good working relation between the participating projects. Detailed material from these workshops is available from the references cited.

The main conclusions from these workshops were as follows:

- The European large Photon / Neutron infrastructures will collaborate on the common development of a federated identity system. This system will be based upon the Umbrella system developed in work package 2 of the EuroFEL ESFRI project.
- In order to avoid any double work, PaNdata/WP3 and CRISP/WP16 will be strongly harmonized. Although not every infrastructure is partner in both projects, the system will be developed as a common activity of all European partners. The agreement between PaNdata/WP3 and CRISP/WP16 is that WP3 will concentrate on the implementation / deployment of Umbrella, whereas CRISP will concentrate on further developments.
- Start of implementation is scheduled for spring 2013.
- For handling the necessary activities two teams (management, technical) are set up. They are organized by PSI.

On one hand, this harmonization requires more synchronization work; on the other hand, any federated system will be accepted by the community only if it is a common action. At the end, only one general solution will be accepted.

3 Implementation

3.1 Pilot installation

An important criterion for the local user offices is how large is the effort at each facility to migrate to the Umbrella system and to allow the identification by the central federated Identity Provider in addition to the local identification. This obviously requires modifications of the local WUOs.

As a first test, the respective modifications have been performed on clones of the running WUO systems at PSI, DESY, and ESRF. A cookbook recipe has been provided by the PSI/Umbrella team and the local modifications have been performed by the local IT responsibles. It turned out that the modifications of these WUOs were so small that they could be performed within only few days. The conclusion is that as designed the resources required for the modifications of the local user office systems are very limited. These tests took place in the beginning of 2012.

3.2 'Friendly user' phase

The next step was to expose the Umbrella system to 'real' users. This was very important, because due to the DIY concept users play within the Umbrella

concept an important role. This phase has been performed in spring 2012 and about two dozen of users all over Europe have been invited to participate. By intention, this test has been performed at that early stage on the project, in order to have enough flexibility to incorporate user feedback.

The 'friendly-user' study has been conducted in the form of a guided online test with several modules distributed over 8 weeks. The titles of the modules were: 'account creation', 'change password', 'retrieve lost password', 'retrieve lost Username', 'DOOR[1] registration', 'DUO[2] registration', 'Umbrella WIKI', 'Umbrella issue tracker'. At the start of each module, the user was informed about the goal, the functionality and the expected action and then requested to perform this action. At the end of the module the user was asked to directly return the feedback into the web form.

In the following list you find a summary of the most important conclusions:

- With the information provided about 90% of the users were able to perform the required actions
- The users pointed out that sufficient information had high priority
- User friendliness was also given high priority.
- Logout handling as realized in the test version was found to be unsatisfactory. In the meantime, this has been improved.

In conclusion, despite the fact, that only the basic functionality was available, the user response was very positive. No problems or even no-go's turned up and the general concept was well accepted.

3.3 Development of a central Umbrella affiliation database

In principle, the hybrid data structure (see above) does not explicitly require that the user information is standardized between the Umbrella partners. If, however, a user should be able to transfer his/her information from one facility to another, a standard is needed. In addition, only a database-structured affiliation definition allows digital handling as e.g. increasingly requested for user statistics for reporting purposes. Furthermore, a common standard will allow synergies, as not each facility needs to administer its own database. In order to guarantee a certain quality level of the affiliation information, users select their affiliation from a list and can propose a new one, if it is not found. The actual input of a new affiliation to the database is performed by staff from the user office.

A special project has therefore been started by the ESRF partner within WP3. The goal is the development of a common European affiliation database, which

[1] DESY local web-based user office system

[2] PSI local web-based user office system

catalogues the affiliations of the users of the PaNdata ODI facilities. As there is no official standard for the definition of address information, the respective entries have to be agreed upon by the Umbrella partners. Consultations have been started and a first workshop was held in Grenoble (October 2012) with the following goals:

- Definition of a common set of mandatory database elements.
- Definition of a common format of these elements (language, specification depth).
- Aliases.
- Development of the respective software modules for manipulation of the database entries and for interfacing the database to the local user office systems.

There was unanimous agreement of the partners that this affiliation database system is needed. The project, however, is too complex for all issues to be solved at one meeting and continuous discussions are necessary. A follow-up TelCo will take place in the coming weeks and this topic will also be part of the implementation-task meetings. A prototype of the systems is planned to be ready in spring 2013, in phase with the general start of the Umbrella implementation / deployment.

4 Next steps

The next tasks concerning the implementation / deployment of Umbrella will be handled within the Harmonization Meetings (mentioned above) and by the implementation teams (management, technical) installed in the 2012 Zürich meeting.

The topics of these meetings are [deadlines indicated in brackets]:

a) Management
   i. Legal issues, development of the MoU between the partner infrastructures [March 2013]
   ii. Definition of the user database (structure and entries) [Dec 2012]
   iii. Definition of the affiliation database entries [Dec 2012]
   iv. Umbrella physical topology [Dec 2012]
   v. Security handling (specification of username/password, security level, handling of incidents) [Dec 2012]
   vi. User support (web info, hotline, trouble handling) [Jan 2013]
   vii. Hardware specifications [Jan 2013]
   viii. Implementation procedure (Common preparations/ First wave participants) [Jan 2013]
   ix. Operation (uptime guarantee/software update handling/user database updating/affiliation database updating) [Jan 2013]

x. Procedure for handling further developments (General procedure definition/specification/coding/tests) [Jan 2013]

b) Technical

i. Development of tools for handling the affiliation database [Apr 2013]
ii. Preparation of options for the management team [on request]

5 Conclusion

The implementation and deployment of the Umbrella system is well on track.
- First, in the beginning of 2012 clones of the standard user-office systems at three partner infrastructures have been updated for the Umbrella service and linked to the central Umbrella server. This took only few days per site, which provides a strong argument for the acceptance of the system at the partner infrastructures.
- Second, from February to April a 'friendly – user' test of the system has been carried through their experience with about two dozen of voluntary users all over Europe. These users were asked to test the system as it is and to feedback. Few detail problems were reported which are in the meantime already resolved. The general response was very positive and confirmative. Specifically, they asked for a system as simple as possible and well documented.
- Third, two teams (management, technical) have been installed with representatives from all partner infrastructures, which meet on a regular basis and take care of all the issues of the implementation / deployment of the system.

6 References

[1] Functional Description of the EUU / UAA Tools, R. Dimper, D. Porte, O. Schwarzkopf, D. Feichtinger, D. Lauk, H.J. Weyer, 2010.

[2] <http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security>

[3] <http://www.switch.aai/index.html>

[4] The Umbrella system Prototype Web-Based Access Point, EuroFEL Deliverable 2.5, B. Abt, Heinz J Weyer, April 2011,

<https://umbrella.psi.ch/euu/documents>

[5] PaNdata / CRISP Harmonization Meeting I, ZRH Airport, Jun 27, 2011 < http://indico.psi.ch/conferenceDisplay.py?confId=1039>

[6] PaNdata / CRISP Harmonization Meeting III, ZRH Airport, Jun 13, 2012

< http://indico.psi.ch/conferenceDisplay.py?confId=1752>

[7] PaNdata / CRISP Harmonization Meeting II, DESY, Dec 8, 2011

< https://indico.desy.de/conferenceDisplay.py?confId=5061>

Appendices

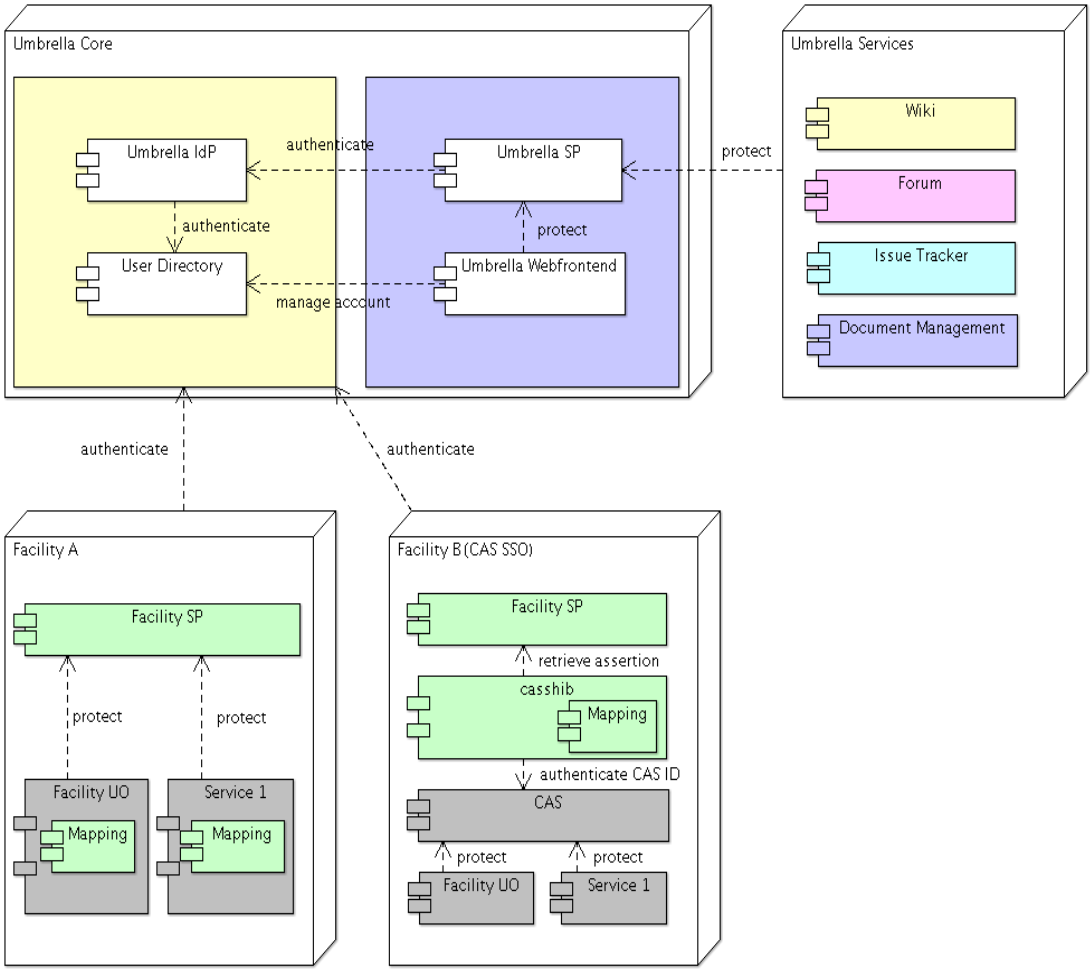
A1 The Umbrella website

Figure 1

A2 Technical description of the Umbrella system

A2.1 Graphical description

Figure 2



A2.2 Explanation of the elements

A2.2.1 Umbrella core

The Umbrella Core provides a central login for all participating facilities.

A User Directory keeps all accounts and provides the functionality to prevent duplicate entries. There is a web front-end (see Fig.1) to enable users to manage their account by themselves.

Umbrella IdP
This is the central authentication endpoint, which allows users to authenticate with their Umbrella account and to consume Umbrella enabled services. It also allows users to logout from all services. The technology behind it is SAML2 and the implementing software is Shibboleth. All Service Providers have a trust with this IdP. It is connected to the User Directory from where authentication is verified and attributes are released.

User Directory
The User Directory holds all Umbrella accounts. It is an LDAP database.

Umbrella SP
The Umbrella Service Provider (SP) uses the SAML2 technology and its function is to protect the Umbrella web front-end and its services from unsolicited access by forcing and Umbrella authentication to protected pages.

Umbrella web front-end
The web front-end is the main management tool for users to administer their Umbrella accounts. Accounts can be created, passwords can be retrieved or changed and affiliation changes, e.g. address changes, can be propagated from here. It is a web application which is connected to the User Directory to perform account modifications and to the Umbrella SP to prevent unsolicited access. (see Fig.1).

A2.2.2 Umbrella services

The Umbrella Services are a bunch of everyday applications using and extending the functionality of the Umbrella. All of these services are shibbolized and therefore allow login with the Umbrella account.

Wiki
A wiki is provided in order to give Umbrella users the possibility to exchange knowledge between other Umbrella users. Umbrella developers can use the wiki to document developments, features and concepts of the Umbrella. For Umbrella MediaWiki is used as the wiki software

Forum
A forum can be used to discuss different aspects of the Umbrella, e.g. 'How to use it', 'How to Integrate an Application', etc. The forum is part of the wiki.

Issue Tracker

The issue tracker is used to report technical deficiencies to the Umbrella technical staff, e.g. 'I can't login', 'My email address has changed. What to do?', etc. Here the open source software Trac is used.

Document Management
The document management is used to store documents and presentations which then can be retrieved on the Umbrella website. Authentication with an Umbrella account is needed to get access to any document. The software used is Alfresco.

A2.2.3 Facility A

Facility A represents a typical facility which has no Single Sign-On system for all their applications. All applications which should participate in Umbrella must be shibbolized.

Facility SP
Each facility needs a Service Provider which has a trust relationship with the Umbrella IdP. It protects the applications from unsolicited access and negotiates the authentication with the Umbrella Identity Provider.

Mapping
A mapping mechanism concept is provided and must be implemented for each application. The goal is to connect the Umbrella account with a local account from an application. This is achieved by letting an Umbrella-authenticated user login to the local application. The Umbrella ID is then stored locally in a mapping table. When the connection between the Umbrella account and the local account has been established, Single Sign-On is possible.

Facility UO
The Facility UO is the facility User Office Software. It is usually used to submit proposals.

Facility Service 1

The facility service 1 is a dummy for a real application hosted by the facility and provided to its users. This application can participate in the Umbrella Single Sign-On system.

A2.2.4 Facility B (CAS SSO)

Facility B represents a facility which already has some kind of Single Sign-On system to protect its applications. In this example CAS is used. This software is already in use at Diamond Light Source and is planned to be used at ILL (Institute Laue-Langevin). Here CAS is just used as an example of such a system to explain the functional pattern. Any other Web SSO could also be used similarly.

Facility SP (ref)
See description in Facility A

casshib

casshib is an open source software which is able to translate authentication between CAS and SAML2. This software is used to retrieve and log in a local user which corresponds to an Umbrella user.

Mapping
See description in Facility A

CAS
CAS, Central Authentication Service, is a single sign-on software package and protocol for the web. It is used in hundreds of university campuses. It was initially developed at Yale University and is now developed and maintained by JASIG (Java Architectures Special Interest Group)

Facility UO
See description in Facility A

Facility Service 1
See description in Facility A