Summary of the 2nd Federated Identity System for Scientific Collaborations Workshop

Philip Kershaw
Centre for Environmental Data Archival
RAL Space
STFC Rutherford Appleton Laboratory
29th November 2011

This second workshop builds directly on the outcomes of the first event held earlier this year at CERN. The goal was to explore the requirements for federated identity management (FIM) drawing input from across a range of different scientific disciplines.   It was agreed that follow on workshops should be organized to advance a roadmap. A number of initial steps were proposed for this process: to identify a small number of use cases from each community represented and to engage with other stakeholders - national and international bodies representing common assets and infrastructures. A team of representatives or architects from each community has been formed to co-ordinate this activity:

- CERN/HEP - Romain Wartel and Dave Kelsey
- Life sciences – Andrew Lyall
- Humanities – Daan Broeder
- Photon/neutron – Heinz J Weyer
- Climate science – Philip Kershaw

The workshop was held over two half-day sessions:

- Day 1: a review of the first workshop followed by presentations from each of the architects on a use case for FIM which needs to be met for their community. The architects presented a collective vision for what they would like to see achieved as a whole for FIM.
- Day 2: representatives from stakeholder bodies make presentations based on a set of questions prepared by the architects. These identify key issues they would like to see addressed in order to realize the vision. The workshop concluded with a review of the vision statement in the light of the stakeholder inputs and agreement on the next steps for the roadmap.

The presentations for the Life Sciences from Andrew Lyall (EBI) and Tommi Nyrönen (CSC) illustrated how there are number of significant challenges which are unique to this community. There is a broad spectrum in the level of security required. For example, the human genome is publicly available yet at the other extreme, personal data must be carefully secured to protect privacy. There is a need for certification, a formal mechanism to confirm a person is who they say they are and also to manage individual’s consent to the use of data about them.

The scope of consent is handled through ethics committees. In addition, services carry with them different requirements for level of assurance, and access attributes and data processing is carried out in mixed environments where there are open and closed datasets. The user base encompasses lawyers, ethicists and policy makers as well as scientists. Service providers are largely unaware of what Identity Providers (IdPs) could do for them to help with the identity management burden. There are around three million users across various biomedical service providers. This makes it difficult to get a picture of individual usage. ELIXIR is seeking to address many of these issues.

Dieter van Uytvanck represented the Humanities community on behalf of Daan Broeder who could not attend. CLARIN illustrates the issues associated with creating a federated system spanning the IdPs of multiple national infrastructures. They have an ambitious use case: a researcher creates a virtual collection of resources from different repositories spanning different organisations where each one requires authorisation (i.e. licences signed). Key barriers to this goal are the varying user attribute release policies amongst IdPs and the need to semantically harmonise user attributes. The difficulties associated with working with only three federations (Finnish, Dutch and German) will be compounded if this is scaled up to all European countries. There are number of other key use cases to note: support for library walk in accounts where users are authenticated in terms of an originating institution rather than an individual identity; support for non-browser based access and for user delegation.

The photon/neutron community has taken a very different approach. Rather than use IdPs from national infrastructures Heinz explained how the Umbrella system uses a single IdP with local sites holding additional information for authorisation. They have a large, mobile user community together with thousands of experiments being executed at the various facilities in the represented. Users typically convene for only a short time period for a given experiment. Consequently standardised or automated management of access rights is needed to manage this burden. Support for a high degree of confidentiality is needed for some user communities particularly structural biology where there is strong competition between groups.

HEP is a large and dispersed community so WLCG provides a good use case since the user base is highly distributed and have very varying needs. A heterogeneous environment of domains, policies and technologies is driving various needs: to define levels of assurance and enforce policy based on these; mechanisms to aggregate attributes from different sources, the adoption of standard namespaces for attributes and credential translation. The role of the IGTF is essential. There is a need for HEP sites to gain accreditation for both X.509 and non X.509 based authentication systems. Romain highlighted that federated identity makes user identities more valuable and hence a high value target for phishing attacks. Further work is needed in this area to protect IdPs and users.

In the climate science community, the Earth System Grid Federation (ESGF) has deployed a complete infrastructure including IdPs independent of existing national providers. In this way it has been possible to tailor to the needs of the community. Nevertheless there is a need to integrate with existing infrastructures and the EGI-INSPIRE project is focused on this goal. In the UK, a new initiative CEMS will leverage ESGF. CEMS, the centre for Climate and Environmental Monitoring from Space is a public private partnership to host services and data to support climate change research and the development of commercial downstream applications. As a collaborative venture FIM will be essential to the infrastructure that is developed. CEMS will exploit cloud technologies. The dynamic and elastic functionality of clouds present particular challenges to address both in trust and confidentiality and the consequent level of security it is possible to achieve.

This workshop saw input from a new community, fusion, with Oliver Hemming presenting from the Culham Centre for Fusion Energy. CCFE have a security infrastructure organized in tiers representing successive higher levels of security. Federated access control would be beneficial for access to the middle tier. Adopting a standard FIM solution is of great interest given the international nature of ITER.

Dave Kelsey presented the common vision gathered from the inputs from the represented communities. The discussion that followed highlighted some of the tensions between the provision of identity management infrastructure and the needs of projects and programmes of the individual scientific communities. This was illustrated in a number of key issues. Firstly, there is a sense of an imbalance in favour of user privacy and confidentiality over usability. Legal and financial risks on the one side mean that any case to improve usability must be made strongly if it is to succeed and result in policy changes. A suggestion was made to have a minimal set of attributes provided by all IdPs as a means of standardisation. There are difficulties with this in terms of finding agreement, how IdPs can get this information and justify *why* the IdPs need these attributes.

A second issue is scalability and how best to exploit existing FIM infrastructure. eduGAIN now has 11 SAML federations. There are 37 countries in eduroam. We can't avoid but link these systems together in the medium-term. Middleware to handle the brokering of credentials between systems will be important in the future. There is a need to integrate the requirements across the research communities and speak with a common voice to present these to stakeholders such as eduroam and eduGAIN. There are many issues to tackle and this underlines the need for a roadmap, and that such a roadmap should identify issues to address in order of priority and what is most easily achievable.

The second day of the workshop took input from organisations responsible for FIM infrastructure be they national, international or pan-European bodies. The presentations addressed the set of questions prepared by the community architects:

1. What is your definition of a community?
2. What must a community do to be recognised?

3. What are the most common use cases you foresee and what is the experience and lessons learnt from supporting such cases?
4. What do you consider to be an acceptable administrative overhead for a user to connect to a service with respect to the actions that the user and actions that the IdP and SP administration have to perform?
5. Are IdPs really in a position to be able to serve many concurrent and distinct projects?
6. Can a per-project release policy be supported and if not what are the alternatives?
7. Access to and definition of identity attributes is a problem facing most communities. Are there any efforts to bring harmonisation of identity attributes across federations?

1) A community can be expressed as a group working together using a common infrastructure. This might be expressed as a VO for example. Interestingly, SURFfederatie deliberately target groups in order to elicit input for innovation. 2) To become recognised, a community must channel effort toward a common goal and be supported by funding. It should, preferably, have a security architect as contact person.

3) Inter-federation is a key use case. There is a need to bridge European federations in collaborations and beyond to global scope. This brings with it the need to bridge independent jurisdictions, licence models and funding programmes. As projects extend beyond existing infrastructures, who has authority, national providers or NRENs? Experience with projects such as CLARIN, illustrates the difficulty of current models scaling to meet such use cases. In addition to projects, users themselves are becoming increasingly mobile moving between institutions and so there is a need to support migration between federations. The IGTF has an important role to play in this whole area.

There are a number of more technology focused use cases. CLARIN is experimenting with user delegation for workflows. Integration across browser and non-browser based user agents was also sited as an important use case.

4) Users should not need to contact Service Providers or their IdP to arrange release policy. 5) There is a need to simplify identity management services and separate this infrastructure from services associated with given projects. 6) There is general agreement that IdPs cannot support per-project or per-Service Provider attribute release policies, as it will not scale. One solution is to aggregate attributes into policy bundles suitable for multiple providers. 7) A number of initiatives are underway in the area of attribute harmonisation including eduGAIN, REFEDS[1] and in the Internet2 community with MACE-DIR[2]. The attribute WG will have recommendations ready by February 2012. The next Federated Identity workshop in Taipei will provide a good opportunity to feed

[1] http://www.terena.org/activities/refeds/

[2] http://middleware.internet2.edu/dir/

into this process. Strong use cases are essential to drive the process of harmonisation forward.

The workshop concluded with a review of the vision statement. A number of additional points were added in the light of the discussions and inputs from the individual speakers. The following next steps have been identified:

- Write-up the common vision as a joint paper by the *architects* from all the user communities with recommendations
- Include in the paper a series of recommendations that will simplify the deployment of the use cases
  - These should be few in number and precise (i.e. what to do and who do we want to do it)
- 3rd workshop is scheduled for 26th Feb 2012 in conjunction with ISGC2012 (Taipei) to engage with Asian colleagues.
-
- Each user community should discuss the paper's contents internally and get it endorse the contents in time for the next workshop.

The humanities community will investigate if they could host the fourth workshop in the summer of 2012.