# Summary of the Federated identity system for scientific collaborations workshop

Bob Jones
IT department, CERN
30th August 2011

This document provides a summary of the workshop entitled "Federated identity system for scientific collaborations" that was held at CERN on 9–10 June 2011.

The idea of organising the workshop came from a discussion with IT leaders of EIROforum[1] laboratories during a meeting hosted by ESA in Frascati, Italy last January. That meeting showed that these laboratories, as well as national and regional research organizations, are facing similar challenges. The user communities served by these organizations are growing. They are also becoming younger and this younger generation, *the Facebook generation*[2], has little tolerance for artificial barriers, many being the relics of technology past supporting policies that could, if reasoned, also evolve. This Facebook generation expects to be able to share data, software, results, thoughts and emotions with whom they choose, when they choose. Their boundaries between work and social life are less sharp, and they expect the tools they use to blend into this environment seamlessly. In addition many of the users have accounts at several research organisations and will need to use services provided by yet more organisations involved in scientific collaborations. All these identities and services need to be able work together without the users' needing to remember a growing number of accounts and passwords. Hence the users have high expectations and while this workshop cannot provide all the answers, it is a significant first step.

The goal of the workshop was to explore the requirements for federated identity management across the different disciplines, compare the functionality, operational constraints and state of deployment of current technologies in order to formulate a roadmap for how we could establish such a service in the future. There were approximately 80 participants from Europe and the USA and the event was broadcast over the internet using EVO. A number of representatives from international organisations in Geneva, such as the Red Cross and Red Crescent, also followed the event online. The participants included representatives from a variety of research communities, many of them linked to the ESFRI[3] supported Research Infrastructure projects, as well as representatives of national and international infrastructures that provide identity related services, standards forums and those whose responsibility it is to decide what identity mechanisms will be recommended to policy making bodies.

All presented material is available online: https://indico.cern.ch/event/129364

[1] http://www.eiroforum.org/ EIROforum is a collaboration between eight European intergovernmental scientific research organisations that are responsible for infrastructures and laboratories: CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL.
[2] http://www.wisegeek.com/what-is-the-facebook-generation.htm
[3] European Strategy Forum on Research Infrastructures
http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

The workshop was split into 3 major sessions. The first was centred on the user community's where each community (European photon/neutron facilities, social science & humanities, WLCG, climatology and bioinformatics) was asked to present themselves and their requirements for identity management as well as the current status of usage. This was followed by session where technology providers and infrastructure representatives presented the set of services currently available and how they see the area developing in the future. On the second day the focus turned to policy questions, national approaches and operational issues in dealing with identity systems for a broad set of user communities. This was followed by a panel session where a number of experts listed what they considered to be the most important points for a roadmap to establish a federated identity management system. The workshop closed with brief summaries of each session by the chairmen and suggestions for possible next steps. More details of each session and the conclusions reached are provided below.

The characteristics and current status of the user communities present were captured in Figure 1 below. The **user community** column identifies the scientific research community represented; the **other projects** column lists those projects which are working with the user community and have an interest in identity management; the **#users** column gives an estimate of the number of individuals in the user community and hence an impression of possible scale of usage; the **chosen technology** column shows which technologies are in use or being considered by the community; the **status** column indicates the level of maturity of identity management deployment; the **IGTF** column indicates if the community is making use of the International Grid Trust Federation's policies, procedures or services.

| user community | other projects | # users | chosen technology | status | IGTF |
|---|---|---|---|---|---|
| **photon/neutron** | EUROFEL, PanData, CRISP | 10,000 | Shibboleth/SAML | Umbrella prototype | no |
| **Social Sciences and Humanities** | DARIAH, CLARIN, CESSDAH, (DASISH) | hundreds now, potential for 10000+ across SSH | Shibboleth/SAML | CLARIN SP federation - will see if they can use eduGAIN | yes |
| **WLCG** | WLCG | 5900 globally | X509 | production | yes |
| **earth sciences** | Earth System Grid Federation, GENESI-DEC, CMIP5, Metafor, IS-ENES | 5000+ for CIMP5 | OpenID, X.509 and SAML | production - earth system grid | not yet but foresee for EGI integration |
| **life sciences** | ELIXIR & potentially 10 BMI ESFRI projects | millions access data via EBI website | no chosen yet | security included in BioMedBridges project workplan | no |

Figure 1 user community characteristics and status

Distilling the requirements presented during the session showed there are many common needs and hence scope for agreement. In particular, commonality is most evident in the following domains:

- Need for Single Sign-On access control
- Ease of use for part-time users (many researchers only access the ICT systems concerned in-frequently or on a part-time basis)
- Controlling access to data sets or repositories is the most foreseen of use of identity management across the user communities
- Support for homeless (i.e. no hosting institute or organisation) users is essential
- A wide range of tools and technologies are already deployed and such divergence will remain
- A smooth transition from existing systems to a federated identity management system is essential

During the session about existing ICT infrastructures, the speakers reviewed what services are currently available and how they are likely to evolve in the future. This included the services provided by the pan-European HTC and HPC e-infrastructures EGI[4], DIESA[5] and PRACE[6] as well as the Open Science Grid[7] in the USA. The EMI[8] project outlined its plans for identity management within its future grid middleware releases. GEANT[9] presented the eduGAIN service for exchanging trustworthy identity related information between partner federations as well as plans for the upcoming Moonshot project. Terena[10] presented its Certificate Service for servers, users and software.

Concluding on this session, the chairman, Romain Wartel, noted that the there is a wide range of services deployed supporting sub-sets of the user community. He highlighted that trust is needed with accreditation being a key aspect and that IGTF is used as the source of trust for many projects. Different levels of assurance are important but non-trivial to implement and traceability is critical. The global reach of a common, federated identity management system is important for many user communities so interoperability across national boundaries is going to be essential.

The second day of the workshop started with a session on policy issues from international (IGTF), national (Netherlands, Switzerland, UK), site-specific (CERN) and operational (EGI security officer) points of view. The presentation showed that federated policies are well established with trust criteria and delegation down to the home institute. Preparing and implementing these policies requires effort and it is important to include proper management of attributes and group membership which may cross pre-defined boundaries. Concentrating all these aspects into a single identity can increase risk if the users become casual about giving their primary credentials to service providers.

The conclusion drawn on the federated identity roadmap by the panel of experts was that trust is the key factor for identity management and the closer the communities are related the higher the level of trust.

Different communities will want their own autonomy so we should be aiming at a number of independent federations capable of interoperating. In doing so we must not forget that individuals have multiple

[4] http://www.egi.eu/
[5] http://www.deisa.eu/
[6] http://www.prace-project.eu/
[7] http://www.opensciencegrid.org/
[8] http://www.eu-emi.eu/
[9] http://www.geant.net/
[10] http://www.terena.org/

identities and people expect to be able to move seamlessly between social networking tools and other tools specific to their work.

We should aim to establish a high-level collaborative policy and not become tied to a single technology. There are many technologies available and this work will be more adoption and deployment of appropriate standards and tools rather than development. The ability to offer scalable solutions and a number of different levels of assurance will ensure that a range of scientific collaborations can benefit from this work.

Scientific collaborations are excellent use cases for inter-federation. It is important to keep AuthN and AuthZ separate since they involve different stakeholders and policy decisions. We should not forget what happens when someone leaves the community so we must understand how can we withdraw identity and the privileges that go with it.

Finally, a set of possible next steps and pending questions were proposed to the participants:

- Each user community should focus on a small number of use cases
  These use cases should include the issues and ideas presented at this workshop and be refined to capture more concrete requirements. This step can be performed in parallel by each of the user communities using the on-going projects as vehicles for gathering the information.

- Nominate architect(s) from each community
  The architects should understand the community they represent and be knowledgeable about the security & identity domains so as to be able to work with their peers from other user communities.

- Simple test case
  Put in place a wiki that would collect all the information gathered and as a means of exploring the issues and steps involved with the goal that everyone interested should be able to access the wiki using their existing identity.

- Involve other stakeholders
  The participants included representatives from research organisations in Europe and the USA but we need to also include Asia and Latin America to support the needs of global user communities. At which point should we involve industrial and funding agencies - now or when the roadmap is ready?

- Build on the workshop to advance the roadmap
  It was agreed that follow-on workshops should be organised to help advance the roadmap. Successive workshops should be hosted by different user communities with the goal of developing a roadmap to which they all can agree. Work will of course need to continue between the workshops if we wish to advance rapidly.