# FAR Project Final Report

Table of Contents

---

**FAR Project Final Report**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>1</td>
</tr>
<tr>
<td>Executive Summary</td>
<td>2</td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
</tr>
<tr>
<td>Aims and Objectives</td>
<td>2</td>
</tr>
</tbody>
</table>

**Methodology**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation</td>
<td>4</td>
</tr>
<tr>
<td>Outputs and Results</td>
<td>5</td>
</tr>
<tr>
<td>Outcomes</td>
<td>6</td>
</tr>
<tr>
<td>Conclusions</td>
<td>6</td>
</tr>
<tr>
<td>Implications</td>
<td>6</td>
</tr>
<tr>
<td>Recommendations</td>
<td>6</td>
</tr>
<tr>
<td>References</td>
<td>7</td>
</tr>
<tr>
<td>Appendixes</td>
<td>7</td>
</tr>
<tr>
<td>Links to Technical Demonstrators Produced by FAR Project</td>
<td>7</td>
</tr>
<tr>
<td>Glossary</td>
<td>7</td>
</tr>
</tbody>
</table>
The Federated Access to Repositories (FAR) project was funded by the JISC under the Repositories and Preservation Programme.

The project was led by the London School of Economics and Political Science. University of Cambridge was the partner institution.

We would like to acknowledge the help and assistance given by our JISC Programme Manager, Neil Jacobs, the Project Director, Jean Sykes, the DSpace and EPrints development communities, and Chris Awre and the RepoMMan project team at Hull [RepoMMan].
Executive Summary

The aims of the FAR Project was to apply Federated Access Management (FAM) principles to the repository environment. The project produced recommendations on the use of attributes for access control decision; demonstrator versions of DSpace and EPrints which show the attribute control in action and a report suggesting how similar changes could be made to Fedora.

Shibbolised versions of DSpace, EPrints and Fedora all existed at the project start but these generally relied heavily on the repository database for decisions about the permissions available to the user (e.g. whether they are an administrator). FAR researched what was required to extend this functionality so that (in principle) any access control decision could be made from the attributes available from the user’s Identity Provider (IdP). The project did this via code review and the consideration of relevant use cases (mainly centring around the requirements of distributed research groups).

Background

A workshop at the OAI5 conference in April 2007 [OAI5] identified a need for improved attribute managed federated access to repositories, and produced a set of requirements. These requirements addressed the use of the Federated Access Management in the context of various use cases for repository access, including open access use cases, the special roles for privileged access to an institutional repository, and content-related access restrictions. Important scenarios identified included that of usage by cross institutional collaborations and virtual organisations (such as many research groups) and that of the use of Federated Access Management (FAM) in conjunction with the embargo function now present in Fedora [Fedora], EPrints [Eprints] and DSpace [DSpace]. Cross-institutional collaboration in research groups has been important for many years, and this is therefore one area in which the work of the Project benefited the JISC community. It is a continuation of the work of JISC in extending Federated Access Management take-up in the UK Higher Education community, so it brought together two of JISC’s strategic themes, eResearch and Federated Access.

Aims and Objectives

The broad aims of the FAR Project was to apply FAM principles to the repository environment. The Federated Access to Repositories (FAR) project took up the list of requirements produced at OAI5 conference in April 2007 and used it to:

- Create recommendations for attributes to describe authorisation to repositories including passing them, where appropriate, to relevant standards groups
  (https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/AttributeUseReport)
- Develop extensions to the repository software (EPrints and DSpace) for use with Shibboleth [Shibboleth] as exemplar FAM middleware (EPrints – not yet complete; DSpace - https://dspace.far-project.lse.ac.uk/casak-shib.html)
- Test existing work integrating Shibboleth and Fedora and recommend to developers what was needed to meet these requirements
  (https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/FARProjectDemonstratorArchitecture)
- Install demonstration repositories and showed how it worked in practice (EPrints – not yet complete; DSpace - https://dspace.far-project.lse.ac.uk/jspui/)
- Feed the modifications and full documentations into appropriate repository software development process ensuring maintenance through future releases of DSpace/EPrints and Shibboleth
- Develop procedures to consider ways of extending federated access management to new functionality introduced into the repository software products following the conclusion of the project (this is addressed by the same report as the first point)

The Project also ensured that the code enabling Shibboleth integration did not become obsolete with updates to either Shibboleth or the repository software, whether EPrints or DSpace; the project developers used the...
most recent versions of each product and liaised with the teams of repository developers and the Internet2
developers to ensure that this was the case as far as possible. The Project maintained contact with national
federations to keep the Project in line with the latest thinking on cross-federation access to resources.
Methodology

The plan for the first phase of the Project was to analyse use cases to describe the access control needed in the real world, in parallel with a review of the code for DSpace and EPrints to see where access control decisions were made by the software. However, the personnel difficulties faced by the project made the full scale use case analysis impossible, while the software analyses indicated that both products had architectures which could be used for generic access control which would not require specific use cases to be tested. The details are given in the attribute use report (https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/AttributeUseReport).

With this analysis, a set of tests were developed to show that the Project had developed a version of EPrints/DSpace which handled generic authorisation using attributes obtained from Shibboleth. (See https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/EPrintsDevelopmentSpecification for the list for EPrints.) The fundamental requirements were:

- There may be group information derived from other sources than HTTP headers (e.g. the repository backend database).
- It must be possible to have dynamic group membership. In other words, existing groups must be re-evaluated when an individual logs into the system. This could be done by using a structure designed for this or similar purposes (such as DSpace's special groups, which cannot be altered through the repository interface and are assigned to users on login) or by using a prefix to distinguish FAM-derived groups (as was done in the EPrints development).
- Configuration must be flexible enough to allow for multiple attributes to be used as sources for group memberships. The attributes involved must be allowed to be multi-valued.
- Configuration must allow flexible manipulation of the attribute values to obtain group names. In practice, this was achieved through the use of regular expressions in the configuration.
- Hard-coded manipulation of attribute values is permitted where there are requirements for the structure of group names in the repository code.

The demonstration installations on the FAR Project server (EPrints – not yet complete; DSpace - https://dspace.far-project.lse.ac.uk/jspui/) were developed to pass the tests listed. (It should be noted that both demonstrators are installed on machines which have temporary self-signed certificates, so access to them requires a user to accept the certificate for their browser session.)

Implementation

The outline plan at the project outset (described in more detail in the project plan at https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/ProjectPlan) was to:

1. Consult the repository community and compile use cases to cover the ways in which access management is applied to repositories.
2. Analyse the software available for DSpace and EPrints to see how best to apply FAM to it to enable these use cases.
3. Develop demonstration versions which could be tested against the use cases, using the most recent releases of Shibboleth, DSpace and EPrints. These demonstration installation are accessible using usernames and passwords publicly listed on the FAR Project WIKI.
4. Refine these demonstrators into releasable software.
5. Work to integrate changes made during development with the main software development teams for DSpace and EPrints.

This plan was severely compromised by personnel problems in the first four months of the nine month project. The table gives details of the project posts and when they were filled; all were intended to be available for the entire project (Nov 2007-Aug 2008).
On a project of only nine months duration, the man-hours lost as a result of these personnel issues had a devastating effect. While it is likely that the project could have absorbed any one of the losses, given the results of the code analyses, three was too much to handle. From the original plan, this required the effective merging of workpackages 2 and 3, and workpackages 3 and 4; and the late delivery of several outputs. The changes also meant that it was harder to fit in working closely with the repository development communities, particularly that around EPrints, and this has affected the project's ability to fully integrate its outputs with their mainstream development timetables.

One lesson which should be learnt from this is connected to the differences between the original proposal and the final version which was funded. The original version of the project was to be a single institution based development project working solely on one software product. The addition of a second partner led to a doubling of assigned management time, but this proved not enough to keep the project on track given the problems which occurred. Doubling the project work leads to more than double the management overhead.

## Outputs and Results

The principal outputs of the FAR project were software patches for DSpace and EPrints which enabled the use of FAM with these repository products. These two patches have broadly similar functionality, and aim to permit repository managers to use FAM at three possible levels:

1. Replace existing authentication with an implicit authentication system such as Shibboleth or a WebISO? product. The advantages of doing this are that the repository will then share single sign on with other resources using the implicit authentication system, while changes to the repository itself are kept to a minimum.
2. Allow attribute information derived from Shibboleth to override and update information held in the repository database about the user. Which attributes are permitted to override which user information in the database will be configurable in detail. The advantage of this to the repository is that the accuracy of the information about the user is guaranteed by the attribute source (usually an institutional directory maintained as part of the institution's user management procedures) and the repository no longer needs to undertake user management of its own.
3. Use attribute information derived from Shibboleth to supply group membership information to determine the access rights a user has in the repository. This could (but may not necessarily) include whether or not a user is an administrator for the repository, or an editor; or it may mean membership of groups with access to restricted sub-collections within the repository (such as research groups which use the repository to store work in progress, the motivating use case for the whole FAR project). While in DSpace it is possible for such groups to be managed entirely from the Identity Provider (creation and membership changes both being possible with attributes), in EPrints groups need to be created in the repository configuration in order for each required group to be used as a role in the repository (see https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/EprintsRestrictAccess).

Installation and configuration of the software patches is fully documented. Other outputs include documentation of authentication and authorisation mechanisms for EPrints, to be integrated into the main EPrints documentation WIKI.
The DSpace code has been made available as a patch to the existing DSpace core. Interest has been shown by core committers, but patches may take as long as three or four years to get into the core. There are currently many important features distributed as patches against the core. Since DSpace has a code-only distribution policy (installers have to compile the core to set up a server), installers will need to install several patches and setting up the FAR patch in addition will cause little extra work. The project team will continue to monitor the process of getting the patch into the core, and aim to ensure visibility for the patch on the DSpace WIKI. Further work may be done to improve configuration, add OpenID? support and add more examples to the documentation.

**Outcomes**

One outcome of the project has been to highlight the role of authorisation in repository installations. This has in the past encountered a certain amount of scepticism from open access advocates, but even the most open access repository will still limit uploads and administrative access. As repositories take on broader roles and FAM becomes more widely used and understood within institutions, federated access to repositories will become more and more important to give the flexibility and integration with other systems that will be required. See for example Knowing Me, Knowing You, a recent blog post by the JISC Access Management Team [JAMTBlog].

**Conclusions**

It is difficult to work out how hard it will be to add FAM support (beyond simply replacing the authentication system) to a complex application such as a repository product. Beyond access to the source code, which is clearly essential, some properties of a software project which make the process simpler include:

- The existing use of groups for authorisation by the application
- A recognition by the software design of the existence of implicit authentication
- Simple installation on a shared webserver using https for secured access
- Pluggable or at least simple architectures for authentication
- Comprehensive documentation covering authentication and authorisation for programmers and administrators
- No requirement that groups/roles etc. need to be listed in configuration or require code modifications to make the repository recognise them (i.e. it is possible to create on the fly groups of users in the repository from attribute values)
- Use of authentication/authorisation management standards such as XACML

From the list of repository products considered by the FAR Project, DSpace satisfies all but the last, Fedora satisfies all, and EPrints satisfies only the first.

This list is likely to be a useful checklist for any application to which a programmer would like to add in depth FAM support, even if it is not a document repository.

**Implications**

JISC are organising a series of workshops on Federated Access to Repositories to follow on from FAR, involving individuals from the UK repository community. The first of these took place on 22 September 2008.

**Recommendations**

The project does not have any recommendations for the teaching, learning, or research communities, and instead would point software developers at the points made in the Conclusions section of the report.
References

[JAMTBlog] JISC Access Management Team Blog entry, Knowing Me, Knowing You (2008-10-02)
http://access.jiscinvolve.org/2008/10/02/knowing-me-knowing-you/

Appendixes

Links to Technical Demonstrators Produced by FAR Project

Test Architecture https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/FARProjectDemonstratorArchitecture
Test Users https://gabriel.lse.ac.uk/twiki/bin/view/Projects/FAR/DocumentationListingTestUsers
DSpace Demonstrator https://dspace.far-project.lse.ac.uk/jspui/
EPrints Demonstrator https://eprints.far-project.lse.ac.uk

Glossary

Attribute An item of information which describes a person, such as their email address or institutional membership.

FAM (Federated Access Management) The concept of division of authentication and authorisation control between Identity Providers and Service Providers.

Identity Provider The part of a FAM system which handles user authentication and obtains attribute information to pass to the Service Provider.

Implicit Authentication A means of authentication to a resource where the resource does not itself obtain the user’s password. Generally the case for FAM or WebISO? mediated sign on.

Service Provider The part of a FAM system which protects a resource, allowing access to users authenticated against trusted Identity Providers. This is separate from the organisation which owns the resource, also often referred to elsewhere as a Service Provider.

-- SimonMcLeish - 17 Feb 2009