



Gilead Project

End Project Report

Nigel Bruce

ISS, University of Leeds, April 2006

In January 2005, the University of Leeds applied for JISC funding to enable us to become an early adopter of Shibboleth technology. The proposal was submitted on behalf of the University through a partnership between Information Systems Services (ISS), the University Library and the Faculty of Biological Sciences. The University is currently engaged in a major initiative which aims to provide simplified or single sign-on capability to a wide range of internal and external information systems. As part of this endeavour the University aims to transition away from a number of existing access management solutions. The University resolved 18 months ago to rationalise the number of directory services on campus and adopted Microsoft's Active Directory (AD) as an institution-wide LDAP service. The University's aim is to reduce the number of username & password databases it has to populate and manage and the adoption of Shibboleth has been identified as an important component in our drive to simplify access to a number of commonly used teaching and research-orientated resources.

The main objectives of the project, which ran from March 2005 – March 2006, were to:

- Build an institutional Shibboleth Identity Provider based on the Athens Identity Manager
- Use this IdP to authenticate access to the Nathan Bodington VLE
- Use Eduserv's Athens-Shibboleth Gateway to authenticate access to Athens-controlled resources
- Use Guanxi derived Shibboleth Origins to test federation arrangements between Leeds and Manchester Universities
- Modify a number of existing resources to act as Shibboleth Targets, namely:
 - British Education Index
 - mvnForum Bulletin Board
 - Scion's Bioinformatics Resources

This document outlines the extent to which the Gilead project met its objectives and the lessons learned.

Creation of an Institutional IdP

Based on the AthensIM, the institutional IdP is now in production status within the University and is a member of the Athens & SDSS Federations and part of a private Federation between the University of Leeds and Penn State University. We chose to use Eduserv's implementation of Shibboleth because one of the primary aims of the project was to use their Athens Shibboleth Gateway and dispense with the need for classic Athens accounts. We believed that if there were any issue with using the AS gateway it would be easier to obtain support and resolutions to problems if we were using the AthensIM. During the project we worked with Eduserv to resolve a number of inter-operability issues with their Identity Manager and developed a custom attribute processor for use with our Bodington SP. In general we were happy with the quality of support we received from Eduserv though it did sometimes seem to depend on whom you spoke to and the advice was not always consistent. In the early days there seemed to be a lot of confusion regarding the requirements for joining their federation and the attributes they wanted sending back. Understandably it did improve as the year progressed as I think they were on a learning curve with us.

We didn't find MATU particularly useful. By the time they were up and running I think that most 'Early Adopters' and gone beyond the level at which they could offer help. Possibly this was because those sites who came forward to be early adopters in the first round already had a reasonably good knowledge of

Shibboleth and its issues. Going forward we intend to move away from using the AthensIM once version 2.0 of the Internet2 implementation is available. This is because we believe that there is little commercial incentive to EduserV in maintaining the development of their IdP. We feel that it will quickly get left behind in the features and that other Universities in the UK will standardise on the Internet2 version.

Users authenticated against our Active Directory

Running AthensIM on Windows platform under IIS 6.0 meant that initially we chose to use Windows Integrated Authentication (Kerberos). This meant that anyone accessing a resource from a PC that was already authenticated to our AD was taken straight into the resource without being further challenged for their credentials. Unfortunately this created problems that users had a different experience depending on whether they were on campus or off. Also not all PCs within the University are members of our AD. Certainly PCs on the Wireless LAN and on the Halls network aren't members. We decided that it would be better if users had a consistent experience regardless of their location so we shifted to running Tomcat and using LDAP authentication to our AD. We were also mindful that in future we might want to run a different implementation of Shibboleth possibly on a non-Windows platform and that we might then have to change our method of authentication which could confuse users.

LURCIS meta-directory being used as our attribute store

LURCIS is an SQL database which is used to generate accounts for all staff and students. As such it is fed from our SAP HR systems and Student Information System (Banner) and therefore contains aggregated information on all members of the University. We made the decision to use this as our Attribute Authority rather than use an LDAP server. This is because we didn't want to extend the schema of our Active Directory with the EduPerson Attribute Class nor have to run a separate LDAP service for this purpose. Using custom attribute processor we are able to retrieve data from LURCIS using SQL queries and format them as the required attribute assertions. Many discussions on Shibboleth seem to assume that it is necessary to use an LDAP directory service as the attribute authority rather than any proprietary relational database. This 'prejudice' is perhaps a reflection of the 'open source, open standards' background of those currently involved in the Shibboleth community which may disappear as Shibboleth becomes more mainstream.

Implemented Yale CAS system for WebSSO

Though we did this we have decided not to take it any further as it is our belief that a WebSSO solution will be bundled with the next major release of Shibboleth from Internet2.

Development of Resource Providers

As part of this project a number of new services were made available as Shibboleth-enabled resource providers. These include:

MvnForum

For the past year, mvnForum has been fully integrated into the WUNLearn (Bodington) VLE. It uses the Guanxi alternative Shibboleth identity and service providers, together with a specially written Java authenticator class, *org.gilead.auth.ShibbolethBodAuthImpl.class*. This class receives username, real name and group membership attributes from WUNLearn via Guanxi and uses them to create user accounts and populate groups on the fly within mvnForum. This setup is now in use in the M.Sc. in Bioinformatics jointly taught by the Universities of Leeds and Manchester. Currently 250 students are using mvnForum from WUNLearn via Guanxi.

Scion

Our original plan was to install a Shibboleth SP at the Scion website. However, this is a commercially hosted site and the hosting company was, unsurprisingly, unwilling to allow the installation of the code on its server. We therefore constructed a replica of the website on a server at Leeds. Cold Spring Harbor Laboratory Press gave us an online version of the textbook 'Bioinformatics, Sequence and Genome Analysis' by David W. Mount. We were able to demonstrate to both Scion and CSHL Press that we could limit access to the material to students in defined cohorts without releasing information that would

compromise their identities. This site is now a resource used by (and limited to) the students on the M.Sc. in Bioinformatics and receives about 150 accesses per day.

British Education Index (BEI)

Objectives met during the course of the project.

- The creation of a Relational Database and its population with existing BEI data.
- The development of an easily extensible Java servlet capable of delivering information over the internet.
- The integration of the servlet with the Apache HTTP Server and Shibboleth
- The investigation of the capacity of Shibboleth to enable the presentation of free or subscription based services

Over 140,000 records were brought together in one database for the first time. Each record is a reference to an educational resource: a paper, thesis or a publication available on-line. These records had been accumulated over a period of 30 years, and were kept in a variety of formats including simple tagged text files, free-text databases (BRS) and Relational Databases. Relational Database technology offered the simplest, and most effective, means of storing the data and enabling its maintenance. The University of Leeds has wide experience of Microsoft SQL/Server - accordingly it was chosen as the database server. A front-end application was constructed using Microsoft Access to enable its maintenance.

The project envisioned delivering information to users of the World Wide Web through a Java servlet. A key aim was to develop a servlet that was extensible, in order to enable future services to be developed with a minimum of effort. Broadly speaking the servlet accepts HTTP requests, digests the parameters which determine a query, executes the query thus retrieving data from the database above, and generates a response delivering the requested information. The servlet has the following key features

- The support of the HTTP GET and POST methods enabling URI and forms based requests
- The capability of executing parameterised queries chosen from a fixed but easily extendable set
- The capability of delivering information in a variety of formats.

The format is determined by a set of templates which have placeholders for data, and the ability of generate 'dependent' queries - the net result is that all of the data that is deemed relevant to the request can be delivered from a fully normalised Relational Database.

The interface has been developed to the point where it is usable as part of a service, but there is a logistical problem in demonstrating it currently because it is Shibboleth protected and access is enabled only by authentication to the Leeds Identity Provider or the BEI Identity Provider. It also has to be borne in mind that the BEI is a commercial service and access to it is not currently freely available.

Apache HTTP Server (with SSL support) was chosen as the Web Server and Jakarta Tomcat as the servlet container - the latter being driven by the former through the Apache-Tomcat-JK Connector.

Although the project was set up to implement a Service Provider, it became apparent that, in practice, the British Education Index would also need to run its own Identity Provider. Whilst most of the use of the system is anticipated to be by people from FE/HE institutions which would be running their own Identity Providers or using the Athens Shibboleth gateway; or Schools and other institutions which would be catered for by the 'BECTA Identity Provider'; they being part of JISC-run federation, use should not be restricted to that community. Accordingly it was decided to implement a Shibboleth Identity Provider for the BEI and the Internet2 implementation of Shibboleth was used for both Service Provider and Identity Provider. Originally, Shibboleth Target version 1.2.1a and Origin version 1.2.1 were implemented. A small federation was set up comprising the target, origin and the Leeds institutional origin. Later with the release of Shibboleth version 1.3, Shibboleth Service Provider 1.3 and Identity Provider 1.3c were implemented.

A key element for a subscription based service is data security. Just because a user is able to authenticate to an Identity Provider should not automatically enable him/her to access the application which queries the database and delivers content. The Service Provider is likely to be part of a large

federation, with many Identity Providers - only some of which should enable access, and even then perhaps only to only a subset of users.

Application level security is enforced through a *security configuration file* conformant to an XML schema. When using Shibboleth, our intention is to allow or deny access depending upon information in the HTTP headers associated with the request, e.g. access may be allowed because the user has authenticated using a particular Identity Provider and/or has specific attributes (set by the Identity Provider's attribute authority). The assumption is that Identity Providers will set attributes defined as part of the EduPerson specification. A complete explanation is deferred (the interested reader can inspect the code) - but generally speaking access is allowed by satisfying *any* or *all* of the conditions specified in one of the configuration file's *shibboleth/headerAccess* elements. Standard regex (regular expression) pattern matching is employed.

In March 2006 a server was purchased and 'built'. It hosts a Web Server, Servlet Container, Shibboleth Service Provider and Identity Provider, an instance of SQL/Server 2005 and the servlet - all referenced above.

A short-term goal is to trial the service with a select group of individuals/organisations with a special interest in the BEI. Most of the prospective people who would trial such a service are within UK/HE institutions.

We are keen to experiment with or exploit other initiatives funded by JISC. e.g. Open URL Resolver functionality might be usefully incorporated into the interface (the current system uses Edina's GetCopy experimentally).

Athens Integration

The original intention of the project was that we would use the Athens-Shibboleth Gateway from the start of the 05/06 Academic session. However given the large number of Athens resources which were not compliant in August 2005 we decided to postpone the go-live date until all the resources the University subscribes to are accessible via Shibboleth. Though many were expected to be compliant by September '05 there were some very important ones missing including ISI Web of Knowledge and Westlaw. The rationale for using Shibboleth was to make it easier for our users but all our project stakeholders agreed that due to the 'cookie problem' to have some resources accessible via Shibboleth and some not would merely confuse our users and make authentication more difficult rather than easier. ISS and the Library agreed that going live wasn't a realistic option until more SPs had re-engineered their websites to work with the Gateway.

Though we had to postpone the go live date for using the Athens-shibboleth gateway we believe that this aspect of the project has been successful. We have found bugs and helped Eduserv to improve their product. We have also demonstrated that though robust it was not yet ready to be used in a production environment due to the lack of coverage.

Since the Gilead project ended, the University Library decided in April 2006 that they will now go live with the use of the Athens – Shibboleth Gateway from September 2006. Most of the remaining non-compliant resources are in the area of Law and Classic Athens accounts will only be generated for students and staff in the School of Law. A method is also being developed by which staff and students can request their own Athens accounts through a web interface in order to reduce the administrative burden on Library staff.

Bodington VLE

The original plan was to authenticate access to Bodington via our IdP from the start of the 05/06 session. Following consultation with stakeholders it was decided in the Summer of '05 that we would not integrate our Bodington VLE with our Shibboleth Origin. After the Gilead project began the University carried out a review of its VLE strategy. The decision made was that Bodington, in its current form, should no longer be viewed as the institutional VLE for planning purposes. The University edict was that institutional development of Bodington was to be 'severely and rigorously constrained' and that the prioritisation of institutional resources must be based on our future VLE rather than any development of Bodington.

However, even if the University hadn't changed its policy on the use of Bodington the first version of Bodington which natively supports its use as a Shib target was version 2.8. This wasn't due to be

released until Christmas 2005 (in fact its release was many months later than this). As we have a policy of not changing the way our users login during the academic sessions this effectively meant that we would need to wait until the summer of 2006 to make any changes.

ISS has also considered the alternative of using Active Directory to provide LDAP authentication to Bodington as an interim measure. However version 2.6 of Bodington is needed for this. This was due out at the end of July but was delayed and in the end there wasn't enough time available to upgrade Bodington to 2.6 and test it before the start of 05/06 academic session.

At the time of writing this report the University is once again considering whether in the light of delays in procuring and commissioning a new VLE it should re-visit the goal of shibb'ing it main Bodington VLE. No decision has been taken on this.

Other Bodington-related work did proceed as described in the project plan. Though the University of Leeds may discontinue the use of Bodington as an institutional VLE it is still being used by us for the joint delivery of the MSc in Bioinformatics with the University of Manchester. The University of Manchester is still committed to using Bodington and Leeds has not ruled out the continued use of Bodington in some areas. The objective of creating a Federation involving both Leeds and Manchester continued and a number of related resources (mvnForum and Scion) were shibb'ed. This linking of Leeds and Manchester's VLEs has been delayed by the ill health of Professor Booth. However, the WUNLearn Bodington VLEs and the Manchester based one are in place and they both have Shibboleth identity providers protecting access to them. Currently only the identity provider at Leeds is operational, but trials with identity providers at both sites will begin soon. Successful trials have also been carried out using the institutional Shibboleth identity provider to authenticate Leeds students taking online modules using the Angel VLE at Penn State.

Conclusion

In conclusion the Gilead project was very successful in producing the Shibboleth protected resources it set out to create. In other areas we were less successful. Some of the goals we aimed to achieve were thwarted by factors outside of our control. These included delays in the release of new versions of Bodington, a strategic decision by the University's Teaching and Learning Board to abandon Bodington and delays by Athens resource providers in making their sites Gateway-compliant. However, though one year later than we originally planned, we are now committed to using the Athens-Shibboleth gateway from Sept 2006. Having one of the UK's largest Universities using this access control method will hopefully give others confidence in its viability and reliability.

Other successful outcomes of the Gilead project of benefit to the UK He community generally are the work with Eduserv to identify and fix bugs in the AthensIM and the resolution of issues stemming from the membership of multiple federations (as a result of joining SDSS and Touchstone).

During the course of the project presentations have been given at national and regional meetings of the Grid community and within the University raising awareness of the changes that are coming and the benefits of using Shibboleth. Going forward, the University is now keen to work with Sheffield University on a pilot to use GridShib within the White Rose Grid. Once more institutions have IdPs within the JISC Federation we are very keen and willing to work with them on collaborative projects leading to service delivery.

*Nigel Bruce
N.Bruce@leeds.ac.uk
April 2006*