JISC DEVELOPMENT PROGRAMMES

SAFARI UKDA:
Shibboleth Authentication for Access to the Resource Infrastructures of the UKDA

Project

Project Acronym	SAFARI UKDA	**Project ID**	
Project Title	Shibboleth Authentication for Access to the Resource Infrastructures of the UKDA		
Start Date	21 March 2005	**End Date**	20 March 2006
Lead Institution	UK Data Archive, University of Essex		
Project Director	Kevin Schürer		
Project Manager & contact details	Karen Dennison, UK Data Archive, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ. Tel: 01206 873574. Email: kdenn@essex.ac.uk		
Partner Institutions	None		
Project Web URL			
Programme Name (and number)	*Core Middleware*		
Programme Manager	Nicole Harris		

Document

Document Title	*Case study (final report)*		
Author(s) & project role	Karen Dennison, Project Manager		
Date	13 March 2006	**Filename**	SAFARI case study_final.doc
URL	*N/A*		
Access	✓ Project and JISC internal	□ General dissemination	

Document History

Version	Date	Comments
Draft	8 March 2006	First draft written by Karen Dennison to be reviewed by project team.
Final	13 March 2006	Second draft written by Karen Dennison and reviewed by project team.

Table of contents

Acknowledgements

SAFARI UKDA was funded by JISC under the *JISC Core Middleware: Infrastructure* programme. The SAFARI UKDA project team would like to thank the project stakeholders for their advice, opinions and feedback. The project team would also like to thank John-Paul Robinson from the University of Alabama at Birmingham for his technical advice and guidance and for allowing the project to use and adapt the VOSP model.

Executive Summary

The central aim of the project was to apply Shibboleth middleware to some of the resources that make use of the Economic and Social Data Service (ESDS)/Census Registration Service (CRS) one-stop registration system (hosted by the UKDA). UKDA is in the distinctive position of hosting a central registration service for a number of JISC-funded MIMAS, EDINA and ESDS services that are made available to all UK HE/FE institutions. The main challenge was to find and implement a technical model that would allow users to authenticate with their identity provider but where access to the protected resource would ultimately be governed by attributes contained in the one-stop registration database based at the UKDA. The system also needed to be able to cater, and provide the flexibility, for future development of additional layers of access control, including agreement to 'special conditions' for access to particular data collections within particular resources. In addition, the system needed to be compatible with, and provide as little disruption as possible to, the services for which the registration system provides access.

The overall aims of the project were to research (in consultation with the project stakeholders and others) the ways in which Shibboleth middleware could be implemented within the one-stop registration system, to identify, implement and test the best possible solution and finally to recommend whether, and how to, roll-out to the wider community. Were a viable solution not found then the project aimed to recommend instead an alternative way forward and describe and discuss the lessons learnt.

Following consultation with Guanxi, MIMAS, EDINA and Shibboleth developers overseas, four possible solutions to the registration system were identified. Two solutions were eliminated, one was considered to be a 'last resort' failsafe solution and one was identified as the best possible solution (the University of Alabama at Birmingham Virtual Organisation Service Provider – VOSP - model). The VOSP model has a centralised attribute repository and the 'Virtual Organisation' acts as both a proxy service provider and a proxy identity provider that sits between the resources and the identity providers. More details on the choice and technicalities of this model can be found in the 'Methodology' section but it is worth highlighting here that the model requires no additional modifications or plug-ins by service or identity providers and is compatible with the services for which the registration system provides access. The model also caters for, and provides the flexibility for future development of, additional layers of access control.

The VOSP model was successfully implemented and Shibboleth middleware applied to the CRS and to the CHCC Collection of Historical Resources. The project established three shibboleth targets (for CRS, CHCC and a proxy SP that formed part of the 'Virtual Organisation'). This was tested internally using test logins set up by MIMAS, where testers posed as users from MIMAS via the SDSS federation WAYF.

As a result of successful implementation and internal testing we plan to begin our live implementation of Shibboleth for all ESDS, CRS and UKDA resources early summer 2006. This work will involve investigation into how to (and whether we should) link the registrations of existing Athens users who move to Shibboleth. We plan extensive user testing July to August 2006 with the aim of completion of the live implementation by September 2006 (but allowing for slippage to the end of 2006). Our Shibboleth implementation will run in parallel to access via Athens during the period of transition and we also aim for the central registration service to be gateway compliant. Subsequent to consultation with the JISC programme manager, it was decided to present the results of the planned extensive user-testing phase in a separate later report to the JISC and not to include a user-testing phase within

the official time frame of the UKDA SAFARI project. This is to ensure that the most useful feedback as possible is gained from the process.

Background

The UKDA provides a central registration service for the following geographically dispersed resources

- ESDS
- ESDS International
- Census Dissemination Unit
- Census Geography Data Unit (UKBORDERS)
- Census Interaction Data Service
- Samples of Anonymised Records
- CHCC Collection of Historical Censuses

Registration is a prerequisite of the majority of data depositors, and many have differing requirements. Registration includes agreement to a legally binding End User Licence (EUL) that outlines the terms and conditions of use of the data (for example to agree to refrain from attempting to identify individuals in the survey data; to refrain from passing the data on to third parties; to correctly cite the data in publications; and so on). Registration also requires users to submit personal information, including name, institution, email and user type (e.g. undergraduate, staff at institute of higher education). Data owners are able to specify who can use their data and for what purpose - different data collections have different conditions of access and the system requires levels of fine-grained access control. Data owners can also specify 'special conditions' relating to use of their data that are additional to those of the EUL. The registration system has been set up so that users can agree to these conditions online. The one-stop system was developed in order to streamline these needs and to ensure that users' access to the data was as trouble-free as possible. It was launched in August 2002 and has increased usage of these resources many times over. Through this work, the UKDA has at its disposal a knowledgeable team who have been working on the system's development since its inception.

The one-stop registration database currently holds over 30,000 users. These users come from multiple disciplines, covering social sciences, science and technology and the arts and humanities. When completed, the evaluation of the system by representatives of such a wide-ranging sample of the JISC's community will provide a valuable insight into both the way in which users interact with shibbolised targets as well as their perspectives on the functionality of the middleware.

Additionally, the technology developed in the course of this project could be utilised by many other federated services that may wish to apply Shibboleth to a model of dispersed services around a central hub. Finally, this project was intended to augment the community's access to high quality datasets by offering an alternative route to the data that is both robust and flexible.

The primary aim of this new work was to make the user's journey from desk to data as streamlined and convenient as possible. Applying Shibboleth middleware to the system would provide the users with an alternative entrance to the one-stop registration service, enhance the existing service and provide a smooth transition to Shibboleth as access via Athens is phased out. It would give the UKDA more control over access to resources and would allow the services it manages to be used within the Athens-Shibboleth gateway.

Aims and Objectives

The key **objectives** originally specified are listed below together with comments in bold regarding whether those objectives were achieved -

- the establishment of Shibboleth resource targets for ESDS, the CRS and the CHCC Collection of Historical Censuses

The project established three shibboleth targets (for CRS, CHCC and a proxy SP that formed part of the 'Virtual Organisation').

- the embedding of these resource targets within the one-stop registration service

This was successfully achieved.

- the investigation and establishment, within the target system, of a transfer mechanism to identify registered users and thus prevent users from having to register more than once

This was successfully achieved. Access to different resources by the same user is governed by the central registration system and therefore registration is required once only.

- the investigation and establishment of a method of target-to-target communication of the details of special conditions to which users have agreed

These will be dealt with using one or more attributes in our registration database (currently dealt with in our existing system that uses Athens by Athens updating the Athens profile).

- the evaluation of the system via user and stakeholder consultation

Stakeholder consultation and feedback has informed the project throughout its life-cycle. An extensive user-testing phase will be achieved July to August 2006 and a separate report on the findings will be submitted to the JISC. It was felt that delaying user-testing to the live implementation phase would allow us to assess a wider number of 'real-life' user and access scenarios with a larger number of resources.

Methodology

1. Requirements

The registration system currently uses Athens for both authentication and authorisation. This has brought with it some advantages (it is an 'off the peg' solution and is used nationally throughout academia) and some disadvantages (the resources for whom it authenticates users have less control over authorisation than would in some cases be desirable; additionally, the profiles used for recording users' registration status may not be over-written and there is usually in delay in the updating of the profiles with respect to agreement to special conditions which causes user confusion and generates user enquiries). The system developed by the SAFARI team addressed some of the gaps within the current set-up, while retaining its robustness.

In order to slot easily into the existing registration system, the new development had to:

- provide communication from the registration database to the Service Providers (SPs) in order to identify each user's registration status prior to their being granted or denied access to the data
- provide a system in which more fine-grained access control may be applied
- be interoperable with systems already established at MIMAS and EDINA

It is worth noting that the external services making use of the registration system only need to know the following, in terms of attributes:

- user's registration status (registered user, expired user etc.)
- user's entitlement to use data covered by special conditions
- a persistent identifier which will link the user to their registration record, in case of breach

In terms of attributes, the registration system only needs the persistent identifier, which must be applied to the user's registration record, as all other details are gathered during the registration process. That said, it may be possible to collect some information (depending on Data Protection issues) from the Identity Provider (IdP) in order to populate certain fields of the registration database, such as user's name, role, department, institution, email address etc. as this would reduce the burden on the user still further.

2. Possible technical solutions

Following consultation with colleagues from Guanxi, MIMAS and EDINA, and with Shibboleth developers and implementers overseas, four solutions to the registration system development were identified:

(i) The UKDA acts as a **'proxy' IdP** as well as a target, authenticating users a second time against its registration database
(ii) The **UKDA is pulled into the primary authentication flow**, with the WAYF located in Essex and registration-related attributes pulled from the registration database
(iii) The UKDA follows the model of the **Virtual Organisation Service Provider** (VOSP) (courtesy of University of Alabama at Birmingham) with its centralised attribute repository
(iv) The UKDA sets up an **external call** from the SPs to its database, outside the Shibboleth protocol

Preliminary details of each of these options appear in the first version of the dynamic project specification document. After careful consideration of each option, it was decided to follow option (iii), the VOSP model, the details of which are provided below. Option (iv) was considered to be a failsafe solution.

3. The VOSP model

UKDA has modified the VOSP model as shown in the diagram below –

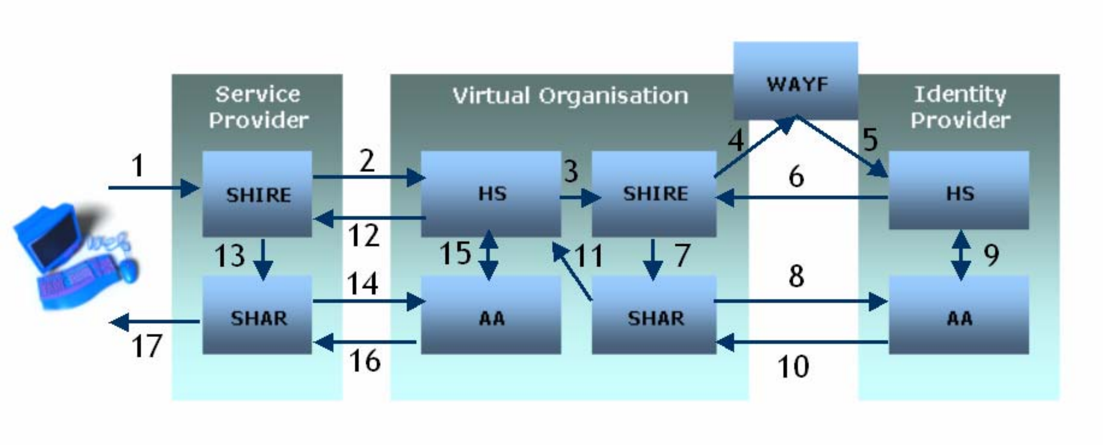


Figure 1: the Virtual Organization Service Provider

In the diagram above, one service provider is shown. This represents one of many service providers for which UKDA provides one-stop registration. SAFARI UKDA is represented by the Virtual Organisation box and consists of a proxy IdP and proxy SP.

The following steps attempt to capture the flow:

(i) a user tries to access the resources of a service provider for which access is governed by the one-stop registration system
(ii) this directs the user to the VO Handle Service (HS)
(iii) a request is sent to the VO SHIRE
(iv) the SHIRE directs the user to the WAYF
(v) the user authenticates at their Home Organisation (HO) (IdP)
(vi) HO HS replies to VO SHIRE with SAML authentication assertion containing a handle
(vii) VO SHIRE hands the handle to the VO SHAR

(viii) VO SHAR uses handle and address of HO AA to request attribute (eduPersonPrincipalName)
(ix) HO AA consults ARP for directory entry corresponding to handle
(x) HO AA releases eduPersonPrincipalName to the VO SHAR
(xi) VO SHAR directed to VO HS
(xii) VO HS directed to service provider (target) SHIRE
(xiii) SP SHIRE passes handle to SP SHAR
(xiv) SP SHAR arrives at VO AA to request attributes
(xv) VO AA consults ARP for directory entry corresponding to handle
(xvi) VO AA releases attributes to SP SHAR
(xvii) based on the attributes, the SP either returns the user to the one-stop registration system (separate SP) or allows access to the protected resource

4. Reason for choosing VOSP model over other models

It was decided not to pursue Options 1 and 2 due to the following issues / obstacles

- Lack of unique identifier within the UKDA AA request
- Requirement for all IdPs to change the data flow by adding a plugin for the UKDA attributes
- Requirement for trust between each Identity Provider (IdP) and the UKDA, as well as trust between each Identity Provider and each Service Provider (SP)
- Requirement to add additional attributes relating to the user's registration status to each resource

The VOSP model was chosen since

- the normal Shibboleth flow is not broken
- access is governed by attributes held in the central registration database (these include attributes indicating whether or not the user is registered and attributes indicating agreement to special conditions). Since UKDA is in control of which attributes are collected, stored and used to govern access this allows for great flexibility with respect to access control and any future development.
- it is possible to employ the scoped eduPersonPrincialName attribute which is persistent across SPs
- there is no requirement for SPs or IdPs to install any additional plugins/make any additional modifications
- no trust issues arise between IdPs and UKDA or between IdPs and other SPs

Implementation

Research was conducted into different technical solutions using advice from stakeholders and others. The best solution was chosen, implemented and tested. Various documents were produced, including a project plan, a progress report and a dynamic specification. Once a technical solution was decided upon and implemented the dynamic specification was updated to reflect this.

Technical implementation
SAFARI UKDA joined the SDSS federation and set up -

- IdP and SP to act as proxy origin and proxy target (VOSP)
- SP called CRS to provide a one-stop registration store where all users need to register. This uses MS sql server database to store user registration information and acts as AA for SAFARI UKDA proxy origin
- SP for CHCC at https://chcc.essex.ac.uk/shibboleth/ registered with SDSS Federation

Outputs and Results

UKDA SAFARI produced a project web site available at http://safari.data-archive.ac.uk/ which includes public versions of some of the project documentation. Also available is a Powerpoint demonstration given at the Core Middleware Programme Meeting (November 2005).

Test web pages were produced to link to the Shibbolized CHCC target (https://chcc.essex.ac.uk/shibboleth/) and to the Shibbolized registration service (https://census.data-archive.ac.uk/shibboleth/). The latter is a copy of the CRS web site that uses Shibboleth and not Athens. The site is currently IP restricted and for this reason some screen shots are provided in the Appendix.

As agreed with the JISC programme manager, the user testing phase that was planned is to be replaced with an extensive user-testing phase July to August 2006 and a separate report on the findings will be submitted to the JISC. It was felt that delaying user-testing to the live implementation phase would allow us to assess a wider number of 'real-life' user and access scenarios with a larger number of resources.

Outcomes

The project achievements largely fulfilled the aims and objectives set (see Aims and Objectives section above). At the outset we were unsure if we would be able to find a viable model to suit our central registration service scenario, so to have found and implemented a model that works and that we will be rolling out to the wider community is the best possible outcome. It was extremely important for us to be able to find a solution. If we had not found a solution then once Athens had been phased out we would have had to abandon access to our one-stop registration system via Shibboleth and reverted to an in-house registration system. This would have inconvenienced users greatly and replacement of the registration system would have been a huge task. Even better, is the fact that we found a solution that suited all our requirements and, although the Shibboleth model we found is not widely used, it falls within the normal Shibboleth flow and requires no additional plug-ins or modifications for UKDA, service providers or identity providers. We hope that other organisations that use a central registration service, or similar, will be able to benefit from the results of this project.

Conclusions

The main conclusion is that Shibboleth can be applied to a central registration service. The VOSP model will allow us to replicate the one-stop registration system using Shibboleth to provide access to multiple resources. In this model the user authenticates at their Home Organisation (IdP) and authorisation is done by the Service Provider based on attributes held in the UKDA registration database (including attributes that are passed to us by the Home Organisation (IdP)).

Implications

Applying Shibboleth middleware to the system will provide our users with an alternative entrance to the one-stop registration service, enhance the existing service and provide a smooth transition to Shibboleth as access via Athens is phased out.

We will need to develop further the work we have done –

- We will need to ensure that, on implementation and once Athens makes it available in several months time, our registration service is compliant with the Athens to Shibboleth Gateway
- During the transition phase from Athens to Shibboleth we will need to design a user friendly interface that allows users to choose whether to log in via Athens or Shibboleth. We will look at what other sites have done to inform us.
- We will need to consider how to deal with a user approaching us using Shibboleth who has previously registered with us using Athens. Specifically, we need to decide whether, and how, we can give the user the option to update their existing account or to link their old account with their new account.

- We need to consider how we are going to deal with users outside UK higher or further education. Currently we have a contract with Eduserv Athens permitting us to issue 15 ids per organisation to users outside UK HE/FE and whose organisations do not issue Athens ids themselves. This could mean that we will issue our own Shibboleth logins and users will then have to authenticate by choosing UK Data Archive as their IdP. A similar system has existed within the SWITCH network. We have also heard that JISC are planning a 'Virtual Home for Identities' for users whose organisations do not belong to a federation.

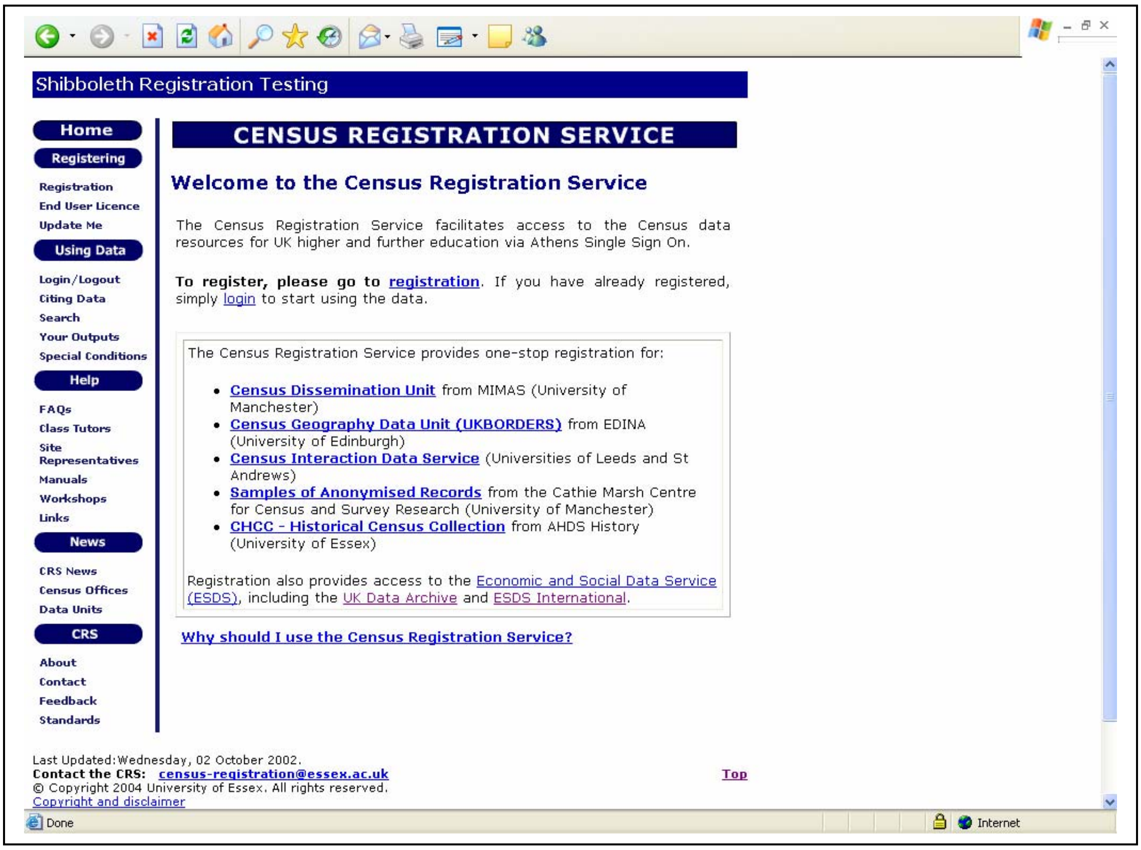
References

http://safari.data-archive.ac.uk/
https://chcc.essex.ac.uk/shibboleth/
https://census.data-archive.ac.uk/shibboleth/

Appendixes

Appendix A

Screen shot of Shibboleth Registration Testing web page



CENSUS REGISTRATION SERVICE

Welcome to the Census Registration Service

The Census Registration Service facilitates access to the Census data resources for UK higher and further education via Athens Single Sign On.

To register, please go to registration. If you have already registered, simply login to start using the data.

The Census Registration Service provides one-stop registration for:

- **Census Dissemination Unit** from MIMAS (University of Manchester)
- **Census Geography Data Unit (UKBORDERS)** from EDINA (University of Edinburgh)
- **Census Interaction Data Service** (Universities of Leeds and St Andrews)
- **Samples of Anonymised Records** from the Cathie Marsh Centre for Census and Survey Research (University of Manchester)
- **CHCC - Historical Census Collection** from AHDS History (University of Essex)

Registration also provides access to the Economic and Social Data Service (ESDS), including the UK Data Archive and ESDS International.

Why should I use the Census Registration Service?

Last Updated:Wednesday, 02 October 2002.
Contact the CRS: census-registration@essex.ac.uk
© Copyright 2004 University of Essex. All rights reserved.
Copyright and disclaimer

Appendix B

Glossary of acronyms

Acronym	**Meaning**
AA	Attribute Authority
ARP	Attribute Release Policy
CHCC	Collection of Historical and Contemporary Census material
CRS	Census Registration Service
ESDS	Economic and Social Data Service
HO	Home Organisation
HS	Handle Service

Acronym	**Meaning**
IdP	Identity Provider
SAFARI UKDA	Shibboleth Authentication For Access to the Research Infrastructures of the UKDA
SHAR	Shibboleth Attribute Requester
SHIRE	Shibboleth Indexing Reference Establisher
SP	Service Provider
UID	Unique Identifier
UKDA	UK Data Archive
VO	Virtual Organisation
VOSP	Virtual Organisation Service Provider
WAYF	Where Are You From service