# Virtual Research Environments: Sakai Demonstrator

Rob Allan, Dharmesh Chohan, Xiao Dong Wang and Xiaobo Yang
e-Science Centre, CCLRC Daresbury Laboratory, UK

Rob Crouchley, Adrian Fish and Miguel Gonzalez
e-Science Centre of Excellence, University of Lancaster, UK

Mark Baker and Hong Ong
Distributed Systems Group, University of Portsmouth, UK

Matthew Dovey and Francisco Pinto
Oxford e-Science Centre, University of Oxford, UK

Corresonding author: r.j.allan@dl.ac.uk

### Abstract

This paper provides an overview of work being undertaken in the JISC-funded Sakai VRE Demonstrator project. This is a collaboration of groups at Daresbury Laboratory and the Universities of Lancaster, Oxford and Portsmouth which are aiming to develop a highly-functional portal interface to distributed resources in a Virtual Research Environment. Sakai was chosen because it already contains a number of collaboration and education tools and has been used to support Grid-based research in the USA, for instance in the NEESGrid project. The paper describes some of the work which is starting to extend the portal framework into a Service Oriented Architecture, in particular authentication and authorisation mechanisms for portal-based VREs, and our approach to evaluating the effectiveness of the deliverables.

## 1 Introduction and Background

E-Science work being undertaken in projects at CCLRC Daresbury Laboratory and the Universities of Lancaster, Oxford and Portsmouth involves investigation and development of advanced portals and tools for e-Research using a range of emerging portlet-based technologies and services via portlet standards such as JSR 168 and WSRP. [1] The Sakai Demonstrator Project is funded by JISC as part of its Virtual Research Environments programme [1] to investigate the effectiveness of a generic portal with a range of collaboration tools and Grid services extending this work for the communities that we support. Sakai was chosen for this on the basis of an objective evaluation exercise [5]. For further information see the Web site http://www.grids.ac.uk/Sakai.

Within a portal, internal services are needed to address the issues of coordination of tools (portlets) within the overall framework. Methods can be provided as an internal class library which sits alongside the portlet API and service APIs (implementing re-usable components of the model part of the MVC design pattern). This is the approach taken in GridSphere (using the GridLab Grid Application Toolkit) for the functionality of its Grid portlets [3, 4] and in Sakai (using the OKI OSIDs [9]) for its educational tools [2]. In the Sakai Demonstrator we are evolving an additional “integration API” with documented extensions to the previous APIs where there are specific requirements for e-Research tasks. The areas we are addressing include: Session management, Authentication using MyProxy, VO management, Integrated state, Service and portlet location, Portal preferences, Semantic/ ontology support, Workflow, Inter portlet communication, and Trails and personalisation.

Each portal framework could have the same or a different set of tools, but the way they are integrated may vary between different application

[1] we will not define commonly-used acronyms in this paper.

areas. Work within JISC on the *Service Oriented Framework for Education and Research* has already classified some services for areas including education, research, collaboration, information access and common services [6]. This is ongoing work and will also explore specifications and standards to enable the implementation of reference models based on these services for a variety of applications. Implementation of the services themselves is currently in a prototype phase. Some simple ones, such as managing the look and feel of the portal, personalisation and accessibility are however provided directly by the portlet framework. Others, such as access to Grid resources, are being provided via a portlet interface to Grid client code using the Java CoG kit [12] or by Web services. This is why the project is currently a demonstrator rather than production software.

In this paper we describe the portal technologies and services that are being developed, extended and implemented in the Sakai Demonstrator. In the first part of the paper we will describe our application areas as well as our original motivation for developing a portal-based Virtual Research Environment. We propose a 3-phase evaluation of the project outcomes. We then discuss the need to extend the basic functionalities currently available via portlets. A more in-depth discussion is given on the use of Shibboleth with a portal-based VRE. In the final part we will summarise our future work, and draw a number of conclusions about using portlets and Grid services to deliver a collaborative e-Research environment.

## 2 Application Areas and support for the Research Community

An important aspect of the programme is to demonstrate potential sustainability and to facilitate different projects sharing tools and service using appropriate definitions and standards. This is naturally evolving in a series of phases which, although not formally defined at the start of the project, have been seen to naturally emerge and will now form part of the evaluation task. These phases are as follows:

**Phase 1:** using Sakai collaboration tools with separate worksites for secure sharing of resources;

**Phase 2:** development and evaluation of portlet-based tools linking to an SOA via a simpler framework;

**Phase 3:** linking portlets and framework extensions into Sakai for fully-functional demonstrator.

Projects currently in the first phase include the following. NWGrid Technical and Operations Boards, for sharing of meeting papers agenda minutes etc. CQeSS: Collaboratory for Quantitative e-Social Science, meeting papers and also collaboration tools for research. They will required Grid-based tools for large-scale modelling via middleware to access applications such as SABRE [13]. ReDRESS: Resource Discovery for Researchers in e-Social Science. The developers and Steering Committee are using separate portal work sites for sharing of papers and collaborative development of content for training and awareness. Eventually the portal-based teaching tools will be used for personal skills development using this content. Sakai Evaluation and VRE are currently using resource management tools to do joint development of papers and presentations. GROWL VRE is using its own work site in a similar way. These projects are typically using the Sakai 'resources', 'discussion', 'e-mail', 'chat' and 'calendar'. We have also found a Wiki tool to be useful and links to the projects' home Web sites and their JISC Mail archives.

Projects in the second phase include NGS, e-HTPX and Integrative Biology. NGS portal: the National Grid Service has deployed a StringBeans container for JSR 168 Grid portlets – see separate paper at this conference. e-HTPX portal: an e-Science Resource for High-Throughput Protein Crystallography has a portal developed using JSP technology with a range of functions for Grid and other resources – see separate paper. IB portal: the Integrtive Biology project has also been funded by JISC to develop a VRE based on the OGCE environment [14]. Initially we have helped them to deploy a portal using GridSphere with the Grid portlets we have developed. This will be used as a platform for demonstrations and extensions to meet scientific the needs of the project. We currently have a suite of portlets including: Single sign-on; MyProxy certificate management; Grid information; Grid FTP; SRB interface; Job submission; Job status monitor.

Work on technology for the third phase has begun and is described in Section 3.

This work is underpinned by a robust deployment of Sakai v2.0 on Oracle 9i. The main portal demonstrator is deployed on an IBM HS20 blade of the e-Science BladeCentre at Daresbury. The Oracle 9i RAC server used is the NGS node at Rutherford Appleton Laboratory. There are instances of Sakai deployed at Lancaster, Oxford and Portsmouth for the development work and we are also investigating mirroring and backup of database contents, multi-site deployment using Apache and WSRP to aggregate portlets across all sites.

## 3 Portal Framework Extensions

In discussion with the Sakai development team, we have identified four generic areas for the demonstrator VRE project to enable a wide selection of tools for collaboration and e-Research to be integrated. Framework extensions are thus being made to accommodate emerging authentication and authorisation systems such as Shibboleth [7] and PERMIS [8] and SOAP-based interaction with remote services such as WS-I Web services, WSRF Grid services and peer-to-peer services. The use of Grid services builds on work carried out in the recent EPSRC-funded OGSA Testbed Project [10]. These extensions will be included in the Integration API and are briefly described as follows.

**Identification, or specification, of an XML grammar for describing of collaborative research.** This work is concerned with the problem of describing a collaborative research session in a standard, easily machine parsable fashion. Firstly we need to be able to describe the start time and duration of the proposed session. We then need to be able to describe its subject matter in as rich a set of terms as possible. This contextual description will be utilised in other tools, e.g. for information retrieval. Finally, we need to be able to describe the participants.

**A Service Authentication and Identity Verification System.** A concern with establishing virtual collaborations, is one of identity. How can you be reasonably sure that the colleague you are working with 500 miles away is the person they say they are? Any Grid tools accessed via Web services are likely to require Grid Security Infrastructure (GSI) type or other appropriate authentication. This work package will implement a Shibboleth Federation [7] consisting of the collaborators' institutions. Figure 3 shows the general security architecture which is further discussed in Section 4.

**Semantic Portal Services.** We wish to enhance the Sakai environment with a number of semantic services to make the portal "smarter" and carry out intelligent reasoning behind the scenes. Adding these services could enable semantics-based browsing and searching, and smart question answering. In the first instance we will develop an advanced portal registry service that will enable the unification of data collected from a range of Sakai data sources, which can be used to support information distribution as well. Associated with this service will be an information manager that will let users create, download, organise, and share any kind technical documentation or information.

We intend to use the RDF model for the information store as it is seen to have a number of advantages over database technologies:

- The RDF model supports extensibility, whereas a relational database has a static schema and it is very difficult to make any

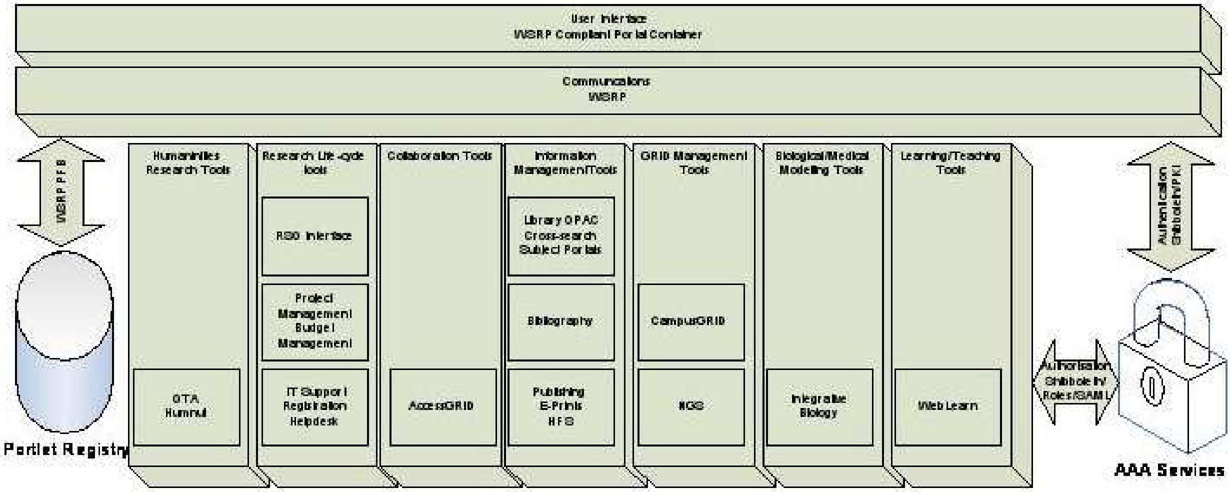
User Interface
WSRP Compliant Portal Container
Communications
WSRP
WSRP PFB
Portlet Registry
Humaninities Research Tools
OTA Humbul
Research Life-cycle tools
RSO Interface
Project Management Budget Management
IT Support Registration Helpdesk
Collaboration Tools
AccessGRID
Information Management Tools
Library OPAC Cross-search Subject Portals
Bibliography
Publishing E-Prints HFS
GRID Management Tools
CampusGRID
NGS
Biological/Medical Modeling Tools
Integrative Biology
Learning/Teaching Tools
WebLearn
Authorisation Shibboleth/ Roles/SAML
Authentication Shibboleth/PKI
AAA Services

Figure 1: VRE Security Architecture

significant changes without impacting existing code. In comparison, RDF model is dynamic and it is easy to add new information.

- An RDF model provides the ability to integrate different resources and there is no limitation on these data sources. A relational database can only store structured data. The RDF model allows structured, semi-structured, and unstructured data can be integrated into the store. So for example a database and web pages can be integrated together.
- The RDF model will aid the development of other higher-level ontological services, which will be critical when automating interactions between the various services integrated within a Portal.

The core of the registry service and information manager will be based around Jena [16], which is a Java framework for building Semantic Web applications. It provides a programmatic environment for RDF, RDFS, and OWL; it includes a rule-based inference engine. We will utilise the Joseki [17] RDF publishing server that provides access to RDF models by URL and query. This will allow individual portals to query and search remote portal registry services and information managers. We will use the emerging RDF Query Language SPARQL [18] that can be used to express queries over RDF graphs. SPARQL defines a protocol for conveying queries, as well as other operations, to an RDF query processing services and conveying the results of such queries and operations to the entity that requested them.

We will also implement a number of other services based on the core services outlined in the previous paragraph, these will include those for translation, annotation, and provenance, these service will be based on work already underway on the Semantic Logging using RDF [19]. The translation service will allow any file or data be uploaded and stored in the RDF store. If the input is already in RDF it will be just incorporated into the graph. If it is not, there is a translation service available that will convert the input into RDF and put it into the RDF store. The annotation service will allow input data to automatically be annotated with metadata that can help with matters such provenance, categorisation, or filtering. The provenance service will help explain how a particular piece of information has been derived. These services are important as the combination of data from diverse sources can result in unique perception difficulties for the information user, these services will provide underlying mechanisms to create semantically rich information.

**A JSF-based Web service interface generator.** If we wish to provide access to Web and Grid services as tools within the Sakai framework, we need to provide a user interface for parameter input. Within Sakai such interfaces are rendered using a pipeline consisting of an abstract XML layout description and a final Java Server Faces user interface. Service can be discovered via registries such as UDDI [11]. A VRE user will be able to search for appropriate Web services to integrate as tools, select the most desirable one and have user interfaces transparently created there and then. Prototype UDDI servers are being hosted at Daresbury, Edinburgh and Oxford through the Grid Engineering Task Force registry group.

**P2P Services and Tools.** We intend to integrate a variety of popular P2P services and tools into the Sakai Framework. The initial phase of work will involve a review and study of the existing Sakai tools and the resolution of those that will benefit from P2P interaction and formulate an appropriate integration strategy that maintains the integrity of Sakai and fits within its security framework. The second phase of the work will involve developing JSR 168 and WSRP components that enable Sakai to benefit from the interaction with external P2P systems outlined in the initial phase of the work package. It is likely that this will include collaborative utilities such as discussion boards, instant messaging, file sharing, calendars, and IRC. The final phase of the work package will create components that will enable interaction with NaradaBrokering [20] and investigate interportlet communications[21]; the former will allow Sakai interact with core e-Science services, and latter has been identified as a missing mechanism from the JSR 168 standard, and will be of general benefit to the community.

# 4 Shibboleth and Portals in the VRE Realm

## 4.1 Shibboleth, Strengths and Limitations

Shibboleth is an identity management system designed to provide federated authentication and authorisation [22]. Once a user has been authenticated, Shibboleth uses an opaque handle to identify the user and exchanges attributes across domains for the primary purpose of authorisation. Its architecture is highly dependent on Public Key Infrastructure (PKI) [23], which is used to build the trust between the several Shibboleth components across the Federation members. Shibboleth is independent of any front-end authentication and authorisation mechanisms. It just specifies how the authentication and authorisation flows should be exchanged securely. End-user authentication based on PKI, and attribute-based authorisation supported by LDAP [24] and eduPerson [25] are suitable, but not a requirement.

In a typical Shibboleth session, users are identified by an Identity Provider (IdP), typically located at their own institutions. Resources and Services are protected by Service Providers (SP), which can be placed within the security domain of any of the Federation members. When a user accesses a protected resource/ service for the first time, the SP needs to find the institution where they are known, which is then used to authenticate the user via Shibboleth's Authentication Authority (aka Handle Service or HS) and later on, during the authorisation phase, to obtain attributes from Shibboleth's Authorisation Authority (aka Attribute Authority or AA). Shibboleth uses a Where Are You From (WAYF) service, which although not mandatory is normally used to determine the institution users belong to or are known by. Having an opaque identifier associated with the user, SPs access the IdP to obtain attributes about her/ him, ensuring anonymity. Users decide which attributes they are happy to release on a resource/ service basis [26].

Portals are normally entry-points to complex information and computational environments requiring the 'identity' of the user to be traveling across multiple tiers in a machine to machine interaction. This requirement creates the need to have entities in the middle of the infrastructure acting as a proxy, posing a security challenge, also know as the n-tier authentication problem [27, 28]. In a typical portal session, a user behind a browser identifies themself to a portal. Beyond that point all accesses to protected resources/ services are 'proxied' on behalf of the user. Credentials are delegated to the n tiers and all the entities involved (including the user) have to trust all components of the chain. Grid environments are not much different of portal environments, if we take out the first 2 tiers (since a Grid session does not require a Web browser).

Shibboleth uses SAML [29] to exchange authentication and authorization assertions. These assertions can be used to pass user identity (or even delegate credentials) and attributes to the n tiers in a trusted and secure way. The current version of Shibboleth (1.3), which implements SAML 1.1 does not address these environments requiring n-tier authentication. SAML 2.0 covers this gap with the specification of the Enhanced Client or Proxy (ECP) Profile [30], which is planned for Shibboleth's 2.0 [31]. Therefore, it will be possible to take advantage of the upcoming ECP Profile to have portals accessing protected data sources on behalf of the user or similarly Grid applications interacting with job managers, and them with protected resources/ services.

### 4.2 Shibboleth and Portal Integration

Shibboleth has been mainly designed for exchanging authorisation attributes across institutions using a secure and devolved model, where users are authenticated and authorised against their own institutional access management system (IdP). Its scope is wider than an institution itself, being very suitable for scenarios known as Virtual Organizations [32] and wider environments such as the JISC Information Environment (IE), where the universe of users does not necessarily belong uniquely to one institution.

Within an institutional context, integration of Shibboleth into a Portal might be done in at least two different approaches: i) natively; ii) integrating it with the institutional authentication system (e.g. WebISO [33] such as CAS [34], Stanford WebAuth [35] or PubCookie [36]) and authorisation system (e.g. role-based authorisation supported by LDAP).

The former approach uses Shibboleth in a WebISO fashion and extends the default mechanisms. Shibboleth is flexible and powerful enough to be used as a native system to provide SSO authentication and authorisation within an institution. However, it is not yet so powerful as some WebISOs, for situations where n-tier authentication (i.e. layered, secondary or proxy authentication) is required, and also situations where authorisation needs to be later assessed as in the case of portals, or non-interactive applications involving Web services or Grid environments.

The latter approach typically uses a WebISO system already deployed and integrated with the institutional Portal and other resources and services. This leaves the WebISO system to provide SSO across the institution for the Portal as if it was any other SSO-enabled Web- based resource/ service. The WebISO authentication point has to be 'Shibbolized' (acting as a SP), at least for all users not coming from a security domain internal to the institution. For users succeeding to authenticate, all the attributes associated with the user might have to be fetched and associated with the WebISO SSO session (since Shibboleth does not go further inside). In the case of a portal, one of the attributes has to be a unique and permanent identifier such as username, e-mail address or something more pseudonymous. We are currentl using the e-mail address.

### 4.3 Discussion

In order to come out with an independent, flexible and staged approach, most institutions are normally interested in the second approach. They deploy locally a WebISO system (instead of Shibboleth), as these systems provide SSO authentication across the institutional security domain. CAS, for instance, is shipped with plugins, configuration files and documentation suitable for a smooth uPortal integration and includes layered authentication. However, these systems generally provide few or no authorisation methods. Therefore, it is the remit of the application(s) to access the attributes exchanged via Shibboleth and collected at the WebISO authentication point or/ and use local authorisation system(s) available (e.g. LDAP), and aggregate the attributes. It is worth mentioning that LDAP could also be used by the WebISO system for the purpose of authentication.

To conclude, a Shibboleth-aware Portal might be the entry point for users seamlessly accessing protected resources and services across different information and computational environments, suitable for inter-institutional collaborations in the context of VREs. The second approach seems suitable and scalable as many institutions have already made significant effort on WebISOs (e.g. WebAuth at Oxford, CAS at Bristol) and the JISC is investing in Shibboleth as the likely next generation access management system for the UK's Higher and Further Education with direct impact in the JISC IE [37].

## 5 Summary and Future Work

Our conclusion from Phase 1 of the evaluation is that members of both the development teams and management of several project have found Sakai to be useful as a collaboration tool. The ability to easily maintain a group of people who can securely share resources has been found to be appealing to project managers and secretaries. This is an outcome which we had not predicted at the start. Just as important as eas-of-use is the fact that Sakai can be maintained as a robust environment for which we have found the fully-supported Oracle implementation on the NGS vital.

For those in Phase 2, the use of Grid based tools is not yet sufficiently advanced to draw conclusions. The ones available so far have generic functionality and need to be augmented by application-specific tools appropriate to the projects' research domains. The fact that tools can be made portable using JSR 168 and WSRP has been shown to be useful so that projects can select from a repository of portlets and augment them in their preferred portal container. There is also a need to make such tools available in a variety of environments, which is why a Service Oriented Architecture is seen to be appropriate with alternative interfaces to programming environments and desktop tools (e.g. GUI based). Links to existing tools and environments, e.g. MicroSoft Office, are clearly important. Standards are the key here to linking between open source and commercial offerings, as was demonstrated using OKI OSIDs and the IMS Global Learning Consortium specifications at the recent international Alt-i-Lab Conference [15] for linking between Sakai and various commercial VLEs.

Phase 3 of the project is ongoing and we are currently exploring the technical issues as described above.

The future goals of this project are to continue the development phases resulting in a fully integrated and highly-functional Sakai portal for e-Research. A report on the evaluation phase will be provided at the end of the project.

## References

[1] *JISC VRE Programme* www.jisc.ac.uk/index.cfm?name=programme_vre

[2] *Sakai Project* http://www.sakaiproject.org

[3] *GridSphere Portal* http://www.gridsphere.org

[4] *Grid Application Toolkit* http://www.gridlab.org/WorkPackages/wp-1/

[5] R. Crouchley, A. Fish, R.J. Allan and D. Chohan *Sakai Evaluation Exercise* Report to JISC (University of Lancaster, December 2004)

[6] R.J. Allan et al. *Open Service Framework Component Classification* www.grids.ac.uk/Papers/Classes/classes.html

[7] *Shibboleth Project* http://shibboleth.internet2.edu/

[8] *PERMIS* http://www.permis.org

[9] *OKI Open Service Interface Definitions* http://www.okiproject.org The Open Knowledge Initiative

[10] Mark Baker et al. *OGSA Testbed Project* www.dsg.port.ac.uk/projects/ogsa-testbed/

[11] R.J. Allan, D. Chohan, X.D. Wang, M. McKeown, J. Colgrave and M. Dovey *UDDI and Web Services Inspection for e-Science* (UK Grid Support Centre, 2002). http://esc.dl.ac.uk/WebServices.

[12] *Globus CoG Kit* http://www.globus.org

[13] D. Stott et al. *SABRE* www.cas.lancs.ac.uk/software/sabre/sabre.html

[14] *OGCE* http://www.collab-ogce.org/nmi/portal

[15] Alt-i-Lab 2005 Advanced Learning Technology Interoperability Conference (Sheffield, 20-23rd June 2005) see http://www.altilab2005.com/ and http://www.imsglobal.org/altilab/. See also www.okiproject.org/project/update_6.html

[16] Jena – A Semantic Web Framework for Java http://jena.sourceforge.net/

[17] Joseki http://www.joseki.org/

[18] SPARQL Protocol for RDF http://www.w3.org/TR/rdf-sparql-protocol/

[19] Semantic Logging using RDF, dsg.port.ac.uk/events/conferences/ccgrid05/wip/schedule/

[20] NaradaBrokering http://www.naradabrokering.org/

[21] JSR 168 FAQ wiki.java.net/bin/view/Portlet/JSR168FAQ

[22] R. L. "Bob" Morgan, S. Cantor, S. Carmody, W. Hoehn and K. Klingenstein. *Federated Security: The Shibboleth Approach*

[23] C. Adams, M. Burmester, Y. Desmedt, M Reiter and P. Zimmermann. *Which PKI (public key infrastructure) is the right one?*

[24] M. Wahl, T. Howes and S. Kille *Lightweight Directory Access Protocol (v3)*

[25] EDUCAUSE/ Internet2 *eduPerson Object Class* http://www.educause.edu/eduperson/

[26] Internet2 *Shibboleth Architecture* shibboleth.internet2.edu/docs/draft-mace-shibboleth-a

[27] Chad La Joie *Trusted Delegation of Privileges in an N-Tier Environment* middleware.internet2.edu/webiso/docs/draft-lajoie-tru

[28] R.L. "Bob" Morgan *Delegation/Intermediaries Use Case Model* staff.washington.edu/rlmorgan/misc/draft-morgan-sstc-

[29] OASIS *Security Assertion Markup Language (SAML)* www.oasis-open.org/committees/tc_home.php?wg_abbrev=s

[30] OASIS *Profiles for the OASIS Security Assertion Markup Language (SAML) V2.0* docs.oasis-open.org/security/saml/v2.0/saml-profiles-

[31] Steven Carmody *Shibboleth 2.0 Roadmap* www.internet2.edu/presentations/spring05/20050503-Shi

[32] A. Saleem, M. Krznari, J. Cohen, S. Newhouse and J. Darlington. *Using the VOM Portal to Manage Policy within Globus Toolkit*

[33] Internet2 *Web Initial Sign-on (WebISO)* http://middleware.internet2.edu/webiso/

[34] *Central Authentication Service* tp.its.yale.edu/tiki/tiki-index.php?page=CentralAuthe

[35] *Stanford WebAuth v3* http://webauthv3.stanford.edu/

[36] *Pubcookie: open-source software for intra-institutional Web authentication* http://www.pubcookie.org/

[37] *Accessing the Future: The Next Generation Access Management System for the UK* www.jisc.ac.uk/index.cfm?name=middleware_futureevent