Virtual Research in the UK: Portal Services for Awareness and Training in e-Science

Rob Crouchley and Adrian Fish
e-Social Science Centre of Excellence, University of Lancaster

Robert Allan and Dharmesh Chohan
e-Science Centre, CCLRC Daresbury Laboratory

Abstract

Staff at University of Lancaster and Daresbury Laboratory are working together on the technical implementation of the ReDReSS portal: *Resource Discovery for Researchers in e-Social Science*. See `http://redress.lancs.ac.uk`. This is an ESRC/ JISC funded project for awareness and training of social scientists to enhance their research through use of e-Science and more general advanced computing methods. Material for the portal, including lectures, seminars, and on-line Grid-based demonstrators are being collected from a variety of sources under the management of a steering committee made up of social scientists with a track record in using computing technology and several advisors. ReDReSS is using the CHEF second generation portal framework. Additional funding was provided by JISC for to extend this work into an evaluation of the new Sakai framework for use in a Virtual Research Environment in the UK.

1 Introduction

In this paper we describe the portal technologies and services that are being developed, extended and implemented to support the e-Science groups with whom we are working. In the first part of the paper we describe our application areas as well as our original motivation for taking up portal technologies; this part of the paper also includes a critique of some of the hurdles that needed to be overcome. We then move on to describe a range of emerging portlet toolkits and services and outline how these can be used to fulfil the needs of our particular applications. In the second part of the paper we present a Services Oriented Architecture which has been developed to extend the chosen second-generation portal architecture to include Web service wrappers of Grid components. This has been created following discussions with Charles Severance and Mark Norton of the Sakai project and with members of the JCSR and JCLT JISC committees. In the third part of the paper we detail our experiences using these emerging technologies; here we also provide examples and discuss implemented services. In the final part of the paper we summarise the paper, and draw a number of conclusions about the using portlets to deliver e-Research and e-Learning collaborative environments and outline our future work plans.

2 Application Areas and Motivation

For the ReDReSS portal, we have chosen initially to use the CHEF collaborative framework [1]. This follows a technical survey by Adrian Fish [2] and conclusions from the international Portals and Portlets 2003 workshop held at NeSC 14-17th July 2003 [3]. In fact the JSR-168 and WSRP standards that we are using were ratified the week following that workshop. CHEF is a content management system aimed at course work in higher education so is ideal for our purposes. Collaboration tools in CHEF, such as calendar, discussion, chat, resource pages and work-group allocation, are already proving useful for developers of ReDReSS and related projects at Daresbury and Lancaster to share material and work on line. We are enhancing CHEF with additional tools, such as ones for presentation of audio-visual lecture material, and a collaborative tool using sound and video for collaborative document preparation and other work involving pairs of people at different sites. The latter is in collaboration with Prof. Geoffrey Fox of University of Indiana and will use the Narada Message Broker [6] to link a portlet to alternative source protocols such as Access Grid, VRVS and H323.

At Daresbury similar portlet technology is being used for other projects with the CHEF framework. The collaboration tools which have been developed in an educational environment are proving to be equally applicable to sharing resources for e-Science Grid-based research projects. CHEF in fact already forms the basis for the OGCE NMI-portal in the USA [8]. This development work will underpin projects such as e-HTPX, e-Minerals, e-Materials and HPCPortal v3.0 for the National Grid Service, NGS, which are being presented in other papers at the All Hands Conference.

3 Second-generation Portal Technologies

The CHEF framework, which is currently being used in ReDReSS, now forms the basis for a much large American project to develop a Collaboration and Learning Environment (CLE) called Sakai [9]. We received additional funding from JISC to evaluate the Sakai framework in the context of a Virtual Research Environment for the UK.

In Sakai, a group of developers from several major research and teaching institutions – University of Michigan and Indiana, MIT and Stanford – are working with the uPortal developers with additional funding from the Mellon and Hewlett Foundations. The project will result in a fully JSR-168 compliant framework with course tools and work tools shared by all the core partners and an additional ten or so institutions in the USA. We are tracking the progress of this project and have had in-depth discussions with the developers about how we may participate. Other JISC-funded evaluations will assess its appropriateness for use in UK HEIs for support of both learning and research. Two workshops were organised by JISC to raise awareness of e-Collaboration which featured presentations from Charles Severance the CHEF/ Sakai chief architect [19].

A the time of writing, the first beta release of the Sakai source and tools has become available as we are attending the Sakai Developers' Conference in Denver 23-25/6/04. Sakai is planned to contain an extensible range of tools and we are exploring how to contribute to this development, in particular accessing distributed services.

4 Services Oriented Architecture

A Services Oriented Architecture (SOA) is an approach to joining up distributed services to provide integrated capabilities. It is a relatively new approach to large-scale software development, but is rapidly gaining popularity because of the lower costs of integration coupled with flexibility and simplified configuration. This is becoming best practice for commercial distributed software development, see recent reviews e.g. [10, 11, 12, 13, 14]. An SOA builds upon the use of Web services, the emerging industry standard for building and integrating distributed systems. Many projects worldwide are considering and indeed beginning to deploy SOAs for distributed applications in both research and learning. One worth noting is Arda, the next generation framework for distributed analysis of Large Hadron Collider data [15]. The UK e-Science community is focussing on Web services for building its open Grid and to participate in the Europe-wide EGEE infrastructure.

Portal deployments have, until now, typically been monolithic with a rich set of locally-deployed tools, customisation possibilities and a database for content management. CHEF, OCE and Sakai fall into this category and are deployed as large Java jar files. At CCLRC the Integrated e-Science Environment, IeSE, comprising services from HPCPortal, InfoPortal and DataPortal has already tried to break this mould [4]. Another more recent activity at University of Indiana is taking a similar approach and now using CHEF [16].

Within a second-generation portal framework there are two complementary ways to leverage the SOA design: one using local portlets which wrap remote Web services; and another which uses WSRP to render remote portlet content locally.

The following figures highlight some basic aspects of an SOA relevant to deploying a Virtual Research Environment with appropriate user interfaces such as portals, online commands, drag and drop desktops and programming libraries. A key aspect of the architecture is to maximise the re-use of common (in many cases already existing) services and middleware including portlets.

Figure 1 shows how an SOA approach would be of benefit in exposing a common set of portlet services and middleware through a variety of user interfaces which include various Web portals employing the WSRP standard. It indicates how this architecture can be used to facilitate the horizontal aggregation that can occur for specific groups, e.g. the National Centre for e-Social Science (NCeSS) which is working alongside the Lancaster node for Quantitative e-Social Science (CQeSS) and the JISC/

ESRC training and awareness programme ReDReSS, see http://redress.lancs.ac.uk. Content can be generated and maintained locally by its authors, but exported for rendering as part of larger resource collections. Discussions have indicated that the Sakai and uPortal frameworks can facilitate exporting an entire workspace from one portal to be inserted in another.

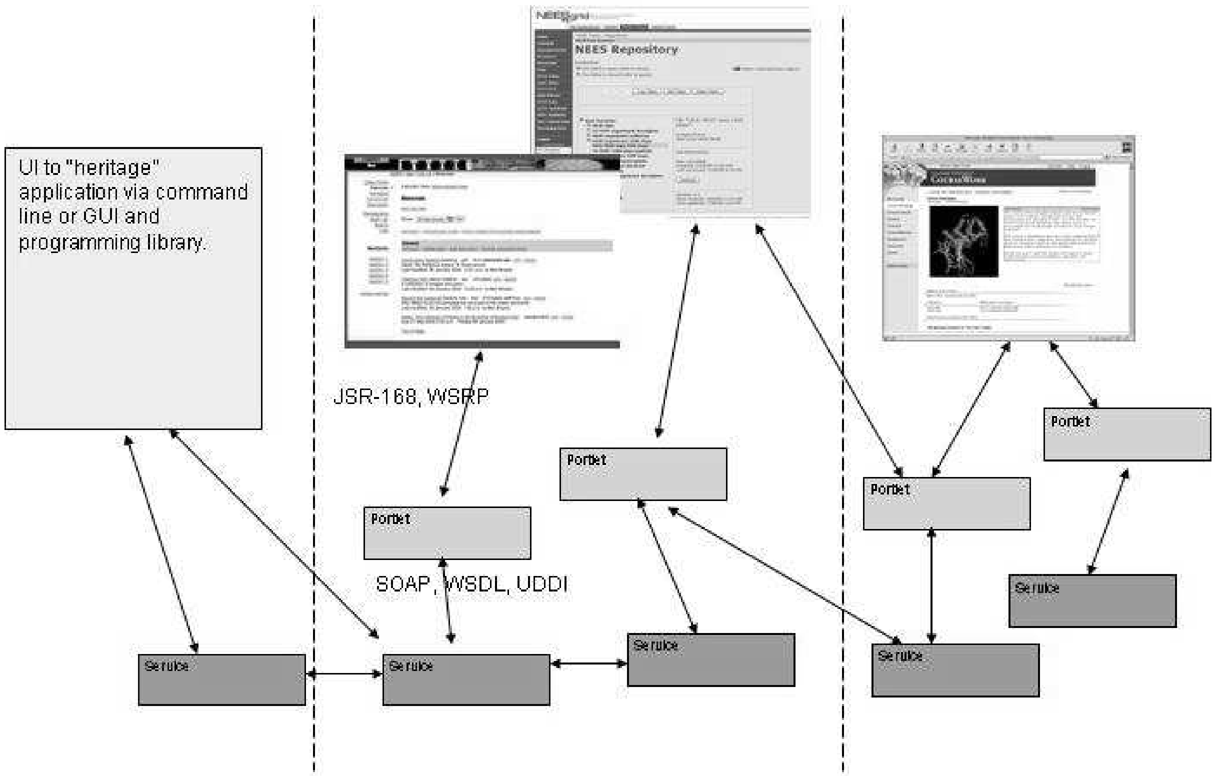

Figure 1: The GRID of Services

This evaluation and consideration of the wider implications and how to implement a Virtual Research Environment for the UK have raised some interesting architectural questions. Similar questions have been raised within the JISC VLE programme, e.g. see the paper by Bill Olivier [18]. The key to both VLE and VRE deployment is to ensure the maximal use of existing resources via a re-usable set of distributed services delivered through a variety of mechanisms such as portals and desktop applications.

In the wider use of SOA involving Web services, it becomes a more complex task to replace the components that provide remote services within the architecture or to look up new ones via a registry such as UDDI [17]. Because service consumers are configured to access a service without any knowledge of the system that provides the service, we can replace the underlying system without affecting systems dependent on its capabilities. Semantic support will however be required to automate this process. This is simple for generic statelss Web services. On the other hand the remote Grid services must have knowledge of the context in which they are invoked, particularly as part of an authentication and authorisation mechanism. This is currently accomplished by passing data objects to service methods and using database structures on the remote server for persistence. Tight coupling into the Sakai framework will be investigated as WSRF implementations evolve.

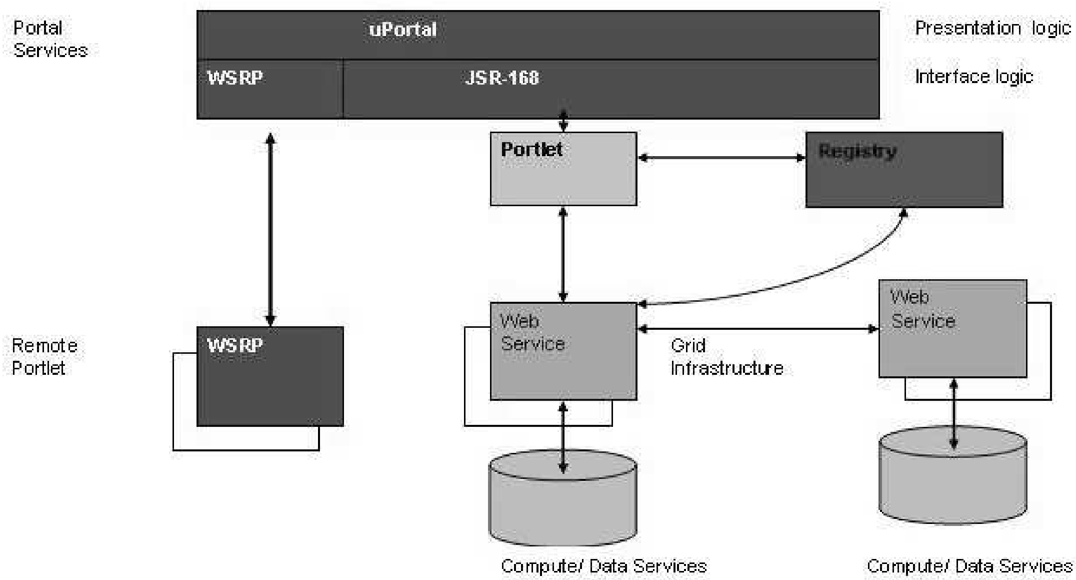

Figure 2: Portlets using Web Services

Figure 2 therefore shows an extension of a portal framework such as uPortal v3.0 to contain a portlet configured to use a service registry. Figure 3 shows how major components in a federated VRE architecture might be linked. In developing this architecture we coined the acronym HIVE: Highly Integrated Virtual Environment. This inherits many aspects from CCLRC's prototype IeSE: Integrated e-Science Environment [4, 5].

In this architecture there can be multiple instances of each component serving slightly different functions. A relatively simple set of basic components might include:

User/ Application: Consumer of delivered services via tools;

Tool Server: User facing part of the system. Browser, programming library, desktop icons etc.

Tool Host: The tools server can be Web or desktop based. Will delegate authentication to HIVE server and thus permit single sign-on across remote toolsets;

HIVE Server: HIVE server provides access to integration services such as authentication, workflow, registries. Can handle federated services;

Shibboleth Server: Will provide the authentication services to the HIVE server. Could be part of a federation and thus provide trust-based access to all the tools hosted for all researchers in the federation's institutions;

VO Management: Provides information about users, their roles and project affiliation. Can extend to resources and services;

Registry: Registry holds details of services and provides template to access them along with relevant semantic information. May be a number

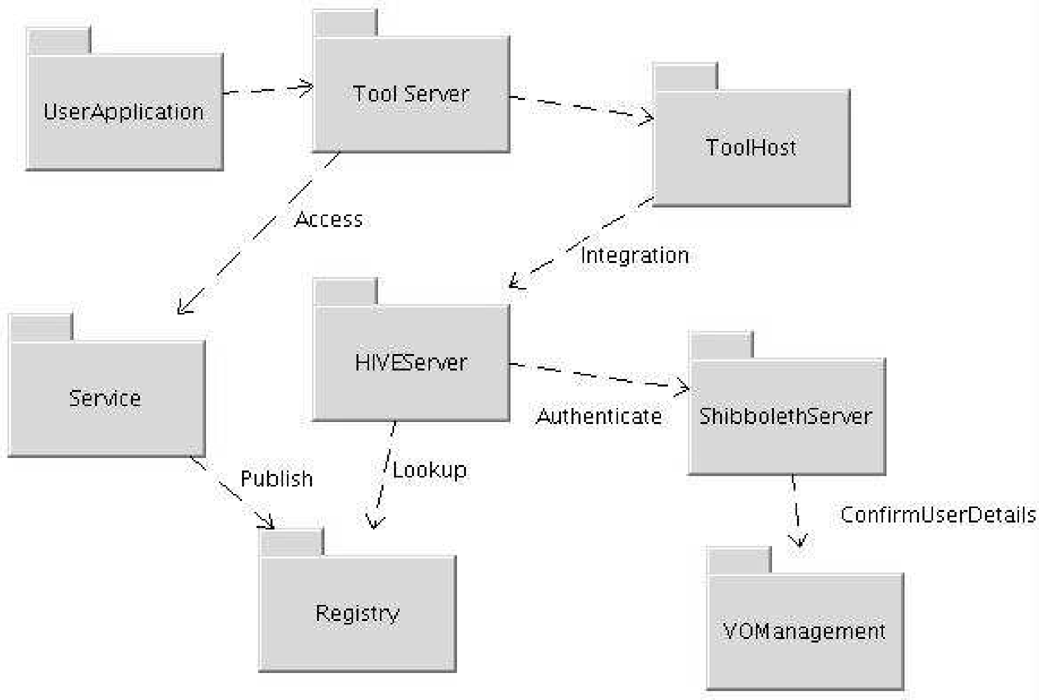

Figure 3: Federated Components in a VRE

of registries handling different types of services. ETF UDDI and JISC IESR are examples;

Services: Multiple services provide access to end resources and applications. Language agnostic so can wrap heritage applications and facilities.

5 Summary and Conclusions

Work on the ReDReSS and other e-Science projects has indicated the need for a small number of tools additional to those already provided in CHEF/ Sakai. These include:

- Authorisation via Shibboleth;
- Context-based information retrieval;
- CopperCore learning content sequencing;
- Creation of application usage documentation;
- Distributed whiteboard;
- Distributed flowchart/ workflow editor;
- Distributed presentation/ authoring tool;
- Distributed display;
- JSF-based Web service interface generator;
- Learning design authoring tool;
- Link to existing UK Grid tools;
- OGSA-DAI portlet;
- Personal and project notebook;
- Personal information management;
- Project bibliography;
- Project specification grammar;
- Research scheduling;
- Scholarly resource cross-searching portlet;
- Semantic common services;
- Service identification and verification;
- Task scheduler;
- Statistical computing Web service interface;
- Video conference and AGN participation;
- Visualisation;
- Wiki

We now report on progress with evaluation and development of a couple of the above tools for Sakai. The outcome of these initial evaluations is promising as we are gaining increasing knoweldge of working with the CHEF framework code base. The release of Sakai 1.0.b2 on 23rd July will enable us to provide a fuller report on this work at the AllHands Conference in September.

5.1 Look and Feel Integration for ReDReSS

CHEF has been reskined to more closely resemble a Lancaster University Web site. This reskinning has been lightweight in that none of the layouts have been changed, only colours and graphics.

5.2 Lancaster AuthN Integration

CHEF's login mechanism has been modified to use Lancaster University's LDAP directory interface for the primary means of authentication with the CHEF internal database-backed system being used as a secondary method. This means that any user, be they student or staff, can log straight into CHEF using their standard network username and password. The piggy backing of an LDAP directory lookup on CHEF's default mechanism like this means that we can also authenticate external ReDReSS users by adding them to the CHEF user database. The CHEF documentation gives some guidance on how to achieve this. This is what we did:

1. Create a java class that implements the interface *PiggybackUdp* You need to implement all of the methods, including the lookup methods, so that they query your institution's LDAP directory implementation. The Udp doesn't just authenticate users, it is used during worksite creation to check that the participants are known to the CHEF software. This gives a degree of referential integrity to the worksite creation process, so this is why the lookup methods are also important. The CHEF documentation recommends that this go in a "plugin" sub-package.
2. Modify the "compile" target in CHEF's `build.xml` file so that the new class is compiled into the CHEF Web application's classtree.
3. Modify the `chef_dev_nc_resources.properties` file in `src/conf` by changing the plugin component class *SampleUserDirectoryProvider* to the fully qualified name of your newly implemented class.

5.3 Lancaster Student Records System Integration

Lancaster uses a students records system called LUSI. This is based on a Microsoft SQL database with a Windows front end for report presentation. It was not possible to gain direct access to this system for the evaluation, so a similar test database has been created. This is linked to a CHEF service called *StudentRecordsService* which has three methods, *getCourseList()*,*getStudentList()* and *getTeacherList()*, the latter two of which take a course id as their sole parameter. This was wired into the Turbine framework using the CHEF configuration files. To permit the adding of a course worksite the *UmiacClient* functionality was replaced with calls to the custom *StudentRecordsService*. This is done, within the *SiteAction* class, by modifying the *finishCourseSite()* and *addNewSite()* methods.

5.4 CopperCore Integration

CHEF has been loosely coupled with the CopperCore IMS Learning Design player from the Open University of the Netherlands. CopperCore is implemented as a set of enterprise Java beans, and as such requires a bean container to run. JBoss is the bean container of choice for the CopperCore team and this is what we installed. We followed the instructions on the CopperCore Web site [20]. Our aim was to enable the attachment of a number of units of learning to a CHEF course worksite for ReDReSS, via a Web-based tool. The mechanisms employed to do this involved JNDI lookups on the JBoss component registry, to retrieve the appropriate CopperCore stub objects; coupled with extensions to the CHEF worksite modification tool *SiteAction.java* that used these stubs to access the CopperCore engine. We modified the CHEF worksite tool by adding a link pointing to a template containing the upload form. We then added a command to the *SiteAction* class to take the uploaded file, save it to disk, and then pass the file to the CopperCore *LDCourseManager* stub for publishing. To enable the transparent addition of all the "access" level CHEF worksite participants to the "learner" role on each unit of learning we had to modify the *LDCourseManagerBean* and its corresponding interface *LDCourseManager* to enable the retrieval of the learner role ids for a supplied Uol (Unit of learning) id. The final stage of the exercise involved the implementation of a worksite tool enumerating the available units of learning and allowing the subsequent playback of the learning material.

5.5 Grid Tools

One aim of the ReDReSS project is to access Grid resources to enable live demonstrations of the sort of application social scientists might want. We are also developing a prototype portal for the National Grid Service (NGS) to provide easy access to resources at CCLRC an the Universities of Leeds, Oxford and Manchester. To enable this we have ported a couple of tools from HPCPortal [5] into the CHEF framework. This was made easier by the fact that the US NMI portal, OGCE [8], is already using CHEF.

A National Grid Service portal is being developed using the CHEF/ Sakai framework for use in a production environment. The NGS portal will consists of the standard CHEF collaboration tools such as Welcome page; Announcements for posting current, time-critical information; Schedule for posting and viewing deadlines, meetings, events, etc; Resources for posting documents, URLs, etc; Discussion for conversation in written form; and Work Site to allow NGS members to setup their own project workspace either as a private Web page or allowing other members to participate in their workspace forming a VO. In addition, Grid tools are being provided to allow users to perform Grid related activities such as Grid Proxy Delegation for creating the x.509 proxy certificate; Grid Resource Broker for job submission; LDAP Browser for resource discovery and GridFTP for file/ data tranfer between the NGS compute and data nodes. examples of these Grid tools had been developed as xportlets by the OGCE team [8, 16]. we have shown that a new Grid tool can be easily integrated into the CHEF/ Sakai portal by adding the tool definition into the XML site description table, adding the tool portlet in the CHEF/ Sakai portlet registry and finally deploying the tool portlet on Apache Tomcat.

6 Generalizations

The HIVE approach to e-Research presented in this paper can be applied in many other contexts. In e-learning we could for example use it to construct a GRID of distributed content that could be dynamically aggregated in the ways required by the user for each learning situation they face. The content of each HIVE instance could be watermarked to identify its origin. This use of the HIVE would require the development new tools, e.g. cross searching tools. Eventually we could use the HIVE to coalesce the appropriate combinations of information, e-Learning, e-Research, e-Collaboration, e-Management, e-Authoring and e-Publishing, e-Leisure tools as required by our current activity.

7 Acknowledgements

We thank JISC and ESRC for funding to carry out the work reported. We also thank the developers of Sakai, in particular Charles Severance, Mark Norton and Glenn Golden for their time to assist in formulating our architectural vision. Other formative discussions have taken place with members of the JCSR VRE working group which resulted in their roadmap document [21].

References

[1] CHEF, `http://www.chefproject.org`

[2] Adrian Fish *ReDReSS Project Review: Survey of Portals and Related Software* (University of Lancaster, February 2004)

[3] Rob Allan, Chris Awre, Mark Baker and Adrian Fish *Portals and Portlets 2003* Proc. NeSC Workshop 14-17/7/2003 `http://www.grids.ac.uk/Papers/Portals/portals.pdf` (CCLRC and NeSC, 2004)

[4] R.J. Allan, D.J. Hanlon, K. Kleese van Dam, S. Sufi, L. Sastry, D.R.S. Boyd and B.M. Matthews *An Integrated e-Science Environment for CLRC: Grid Portals, HPC Computation, Data Management and Advanced Visualisation* Project Overview (CLRC, 2002), `http://www.dl.ac.uk/TCSC/UKHEC/GridWG/`

R.J. Allan, K. Kleese van Dam, L. Sastry, and B.M. Matthews *An Integrated e-Science Environment for CLRC: Grid Portals, HPC Computation, Data Management and Advanced Visualisation* Project Definition, Use Cases and Component Specifications (CLRC, 2002),

[5] A.J. Richards, R.J. Allan, G. Drinkwater and K. Kleese *Advanced Grid Computing* Proc. AHM'2003 (Nottingham, September 2003)

[6] Narada Brokering, `http://www.naradabrokering.org`

[7] Apache Jetspeed, http://jakarta.apache.org/jetspeed

[8] OGCE NMI portal, http://www.ogce.org/index.php

[9] Sakai, http://www.sakaiproject.org

[10] Michael Stevens *The Benefits of a Service-Oriented Architecture* Developer.com http://www.developer.com/services/article.php/1041191

[11] Michael Stevens *Service-Oriented Architecture Introduction* (2 parts) Developer.com http://www.developer.com/services/article.php/1010451

[12] *Web Services and Service-Oriented Architectures* http://www.service-architecture.com/

[13] Sayed Hashimi *Service-Oriented Architecture Explained* (O'Reilly) http://www.ondotnet.com/pub/a/dotnet/2003/08/18/soa_explained.html

[14] Bill Ruh *Succeeding at Service Oriented Architecture* (ZDNet) http://www.zdnet.com.au/builder/architect/work/story/0,2000034884,20276810,00.htm

[15] *The Arda Project – a Realisation of Distributed Analysis for LHC*. http://lcg.web.cern.ch/LCG/peb/arda/Default.htm

[16] D. Gannon et al. *Xportlets Architecture* (Xtreme Computing Laboratory, University of Indiana, 1/4/2004)

[17] Registry interest area of UK Grid Engineering Task Force

http://www.grids.ac.uk/ETF/public/registry/index.shtml

[18] Bill Olivier *Information Paper for JISC VLE Programme* (JISC, June 2004)

[19] *Building Collaborative eResearch Environments* Final report compiled for JISC by Andrew Cox (University of Loughborough, March 2004)

[20] CopperCore, http://www.coppercore.org.

[21] Rob Allan and Rob Crouchley *Roadmap for a UK Virtual Research Environment* JCSR(04)14. Report of the JISC JCSR VRE Working Group (JISC 2004)