A Web/Grid Services Approach for a Virtual Research Environment Implementation

Y.W. Sim[1], C. Wang[1], L.A. Carr[1], H.C. Davies[1], L. Gilbert[1], S. Grange[2], D.E. Millard[1] and G.B. Wills[1]

1. Electronics and Computer Science, University of Southampton, UK
{yws01, cw2, lac, hcd, lg3, dem, gbw}@ecs.soton.ac.uk

2. Royal College of Surgeons of England, London, UK
simon@grange.me.uk

Abstract

The role of scientific and scholarly publications is changing due to advances in Web technology. In particular, the need for collaborative science has generated demands for working environments that facilitate human communication and resource sharing among research communities. Such environments have been typically implemented as monolithic systems in the past, which are then faced with challenges in adapting to changing user requirements. CORE is a JISC funded project to develop a Virtual Research Environment (VRE) that enables orthopaedic surgeons to collaborate in the design, analysis, and dissemination of experiments. The VRE presented in this paper is designed based on the concept of a loosely coupled Service-Oriented Architecture (SOA). This paper reports the user requirements of the VRE and discusses the advantages that will be achieved by implementing the VRE as Web/Grid services.

1. Introduction

Until recently, scientific findings have been routinely captured, summarised, and shared through manuscripts. Nevertheless, the practices of science have already been affected dramatically by the Web technologies initiated at CERN to facilitate rapid information sharing between physicists working in different universities and institutes all over the world. These technologies have also given publishers a new medium for making their journals available [8].

In addition, the rapidly expanding computing and storage capabilities of the federated Grid, coupled with advances in optical networks are accelerating the trends of disseminating and sharing of scientific findings over the Web. However, the current vision for the Grid focuses only upon the immediate aspects of e-science: the experiments, analyses, and meetings that occur over the duration of a project. As well as these synchronic aspects, any scientific effort (and e-Scientific effort in particular) will have diachronic features, collaborative activities which extend through time, enabling the influence of the project to carry on beyond its funded timescale and disseminating its knowledge beyond the boundaries of the original collaboration. These activities are a well known part of the scientist's profession, i.e. publishing papers and data, giving seminars, conducting experiments and verifying results from other research activities, comparing approaches from different projects, generalising or specialising the work of others, and, of course, teaching.

The Collaborative Orthopaedics Research Environment (CORE) [5] is a JISC funded project, which aims to provide an infrastructure that combines clinical, educational and research activities and resources in one working environment. The current paradigm of information sharing and resource use in biology and medicine are being challenged on several fronts. Firstly, as the number of investigators, organisations and institutions conducting biomedical research increases, it becomes difficult to track the work and provide infrastructure to support these expansions. Although current information technology supports ready access, it does not address abstraction, integration and interpretation of information. The diverse bio-informatics tools generated to consume and evaluate the data rarely interoperate [3]. Secondly, the very large volume of data generated in modern biomedicine presents a primary challenge to the researcher. To integrate biological data one would want to move seamlessly between biologic and chemical process, organelle, cell, organ, organ system, individual, family, community and population. Such integration

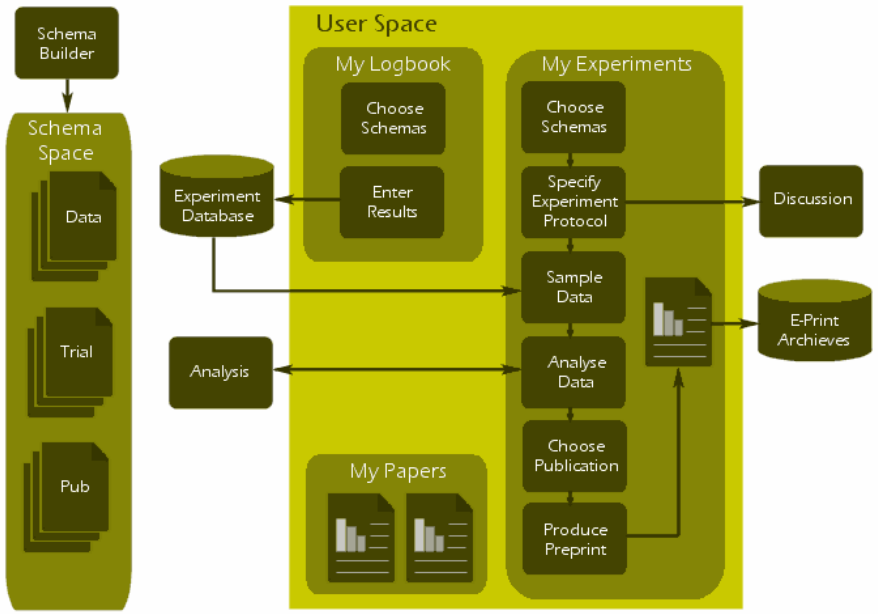

Figure 1: Activities and workflow in the VRE framework

generates challenges to information structure as each research community tends to speak its own scientific dialect [7]. Finally, biomedicine's culture is at the nexus of challenge faced by many other scientific fields: the need for collaborative research. The biomedical researchers recognise that many of the technology approaches required in biology and medicine are expensive, beyond the reach of individual investigators, and increasingly challenge the resource reserves of all but a few institutions.

2. The VRE in VOEU

In response to the challenges discussed above, the CORE project will implement a Virtual Research Environment (VRE) demonstrator using a Service-Oriented Architecture (SOA). The CORE is a follow-up project to the EU-funded Virtual University for Orthopaedics (VOEU) [14], which had two main functions: to support the educational process and to aid surgeons in preparing findings for publication. Orthopaedic surgical trials typically run for extended periods (up to two years), with postoperative assessment results being collected regularly. The collated results are then analysed and discussed by a team of surgeons before being disseminated to the wider orthopaedic community. The activities supported within the VOEU VRE [4][6] (see Figure 1) is described in the following sections before moving on to discuss the CORE approach in re-developing the VRE using a loosely coupled architecture.

2.1 Schema Space

The schema space is the mechanism by which the VRE is configured to a particular e-research community, through the formal specification of e-experimentation procedures relevant to that community. This configuration is achieved using three different types of schema, namely *data schemas*, *experiment schemas* and *publication schemas*.

Data schemas describe the exact nature of the experimental data (for example, specification of variable names, types, and possible values). There are a number of *data schemas* created in the VRE for collecting orthopaedic clinical trial data.

Experiment schemas lay out the procedures or *protocols* for experiments. For example, a protocol could specify that any e-scientist conducting an experiment of type *X* needs to record an *experiment description*, a *statement of purpose* and an *outcome hypothesis*. Human-readable guidelines are also included, to help users meet the requirements of the protocol and to help reviewers to ensure that the requirements have been met. Context sensitive help is provided depending on the role of the user in order to help them understand what should be in any dialogue box. As a simple example, the user guidelines for the *experiment description* may state "summarise the content of the experiment", whereas the reviewer sees "does the experiment description adequately summarise the content of the experiment?"

Publication schemas specify the required format for submitting experimental results to

relevant journals/conferences (for example: *Abstract*, *Introduction*, *Background*, *Experimental Methods*, *Results*, and *Conclusions*). As with *experiment schemas*, human-readable guidelines are also included in *publication schemas*. In VOEU there are currently two *publication schemas* presenting the submission formats for the *Journal of Bone & Joint Surgery* (JBJS) and the *British Medical Journal* (BMJ). Where possible, the *publication schema* also describes any mappings to the experiment protocol, for example, specifying that the experiment *hypothesis* should appear in the *Experimental Methods* section of the article. This allows outline preprint 'previews' to be generated automatically without requiring the user to copy and paste information between protocol and preprint.

2.2 User Space

The user space is where surgeons use the schema space to orchestrate practical data entry and collation, e-experimentation, and dissemination. The user space is further subdivided into three personalised areas *My Logbook*, *My Experiments*, and *My Papers*.

My Logbook (or e-portfolio) is an experiment logbook, in which experimental results can be entered in accordance with a selected data schema. Logbook entries are subsequently added to the VRE community database, making data available (anonymously) to other community members.

My Experiments is a workspace for e-experiments, which the e-scientist works on. An e-scientist may be involved in an experiment in the capacity of *lead investigator* (initiates experiment and acts as co-ordinator and contact for duration of experiment), *associate investigator* (assistant), or *reviewer* (monitors the progress of the experiment and reviews its outcomes according to guidelines). Reviewers have read-only access to the experiment protocol and set-up. When a new experiment is initiated, a discussion facility is automatically set up to facilitate and record communication between the e-scientists involved, and this is also the means by which reviewers can give feedback to the practitioners.

My Paper provides facilities for the users to create preprints. For example, when a user selects a publication schema to create a paper, the VRE toolkit generates a template using the trial protocol information entered by the user. Upon completing the paper, the user can submit it to the community EPrints server, where the paper becomes available to the members of the community.

3. The CORE Approach

The VRE in VOEU is based on an integrated and tightly coupled architecture, making it difficult to expand as the user requirements change. The CORE is a VRE which takes the foundations established in VOEU and decomposes them into modules. These modules will then be developed using Grid/Web services technologies within a SOA, where flexible granular functional components expose service behaviours accessible to other applications via loosely coupled standards-based interfaces. The term, SOA, refers to systems structured as networks of loosely coupled, communicating services [2]. It is a style of design that guides all aspects of creating and using services through their lifecycle (from conception to retirement), as well as defining and providing the information infrastructure that allows different applications to exchange data regardless of the operating systems or programming languages underlying those applications.

Grid services were originally conceived as a method to support virtual communities through sharing of computational and data resources. Access and identity control are fundamental components of the Grid technology, which supports deterministic queries across a distributed, common schema. The technology also supports stateful processes important to the concept of workflow. The developing *Open Grid Service Architecture-Data Access Integration* framework holds promise for adding semantics to the Grid technology so that computable, semantic interoperability may be achieved.

Grid technology does have its limitations. Despite its developing research maturity [9], Grid is a distant second in commercial application. Web services technology is the preferred choice for the vast majority of information infrastructure support installations, in part because of the greater relative simplicity of the technology. Web services extend the concept of SOA into a vast networking platform that allows the publication, deployment, and discovery of service applications on the scale of the Internet. It is a straight forward extension of Internet and Web infrastructure familiar to the majority of systems designers and administrators.

However, the security and state awareness of Grid technology is essential for VRE applications. Hence, the CORE will attempt to combine the Grid and Web services technologies under one architecture, i.e. SOA,

in order to realise the benefits associated with both technologies.

Wilson *et al.* [15] discuss in detail the advantages of using SOA, and in the CORE project's context the following particularly apply:

- **Modularity**: Due to the nature of loose coupling in SOA, applications can be developed and deployed incrementally. Often, a reasonable subset of the full functionality can be developed quickly, which has obvious time-to-deployment advantages. Additional functionality can readily be added in planned stages until the full feature set has been realised.
- **Interoperability**: As the services specifications defined by standard bodies, such as W3C [2] and OASIS [10], progress toward standardisation, it becomes increasingly easy to incorporate services from third parties into the VRE as required.
- **Extensibility**: Some developers will find the core services specification sufficient, while other may not. Hence, the core specifications were defined with built-in extensibility points such as security and reliability of services. Such feature will steer the developers away from the danger of technology lock-in.

3.1 The User Requirements

The end users of the generic toolkit component in CORE will be Higher Surgical Trainees, who are qualified surgeons training to be consultants. Their study is work-based, and they are rarely co-located with other trainees. During the six years of training, these trainees are usually re-located twelve times. In addition, they have to keep an e-Portfolio (logbook) as required by their learning agreements. Therefore, they typify both scientists working on collaborative project, and a group of e-learners studying in a co-partnership institution. An objective of the project is to provide integrated computer support across the research and educational cycles because these activities are intrinsically coupled as part of the requirements of a surgeon's continuing professional development. They must undertake research and publish papers to achieve goals under the learning agreements with their Professional Colleges.

In order to design a VRE with enhanced features that will tailor to end users' needs, the authors have conducted a user requirement study with various experts and users [13]. The results from the study have reinforced the necessity of developing a Grid/Web services-based research environment with the ability to evolve with the changing requirements of its user community. For example, there are vast numbers of protocols to choose from when conducting a trial. Although it might be possible to make a small number of generic trial protocols or templates, it is unreasonable to predefine and encode every possible experimental procedure in most occasions. Likewise, it is also difficult to predict every type of data that can be collated and analysed, and every possible dissemination route that users will follow to publish their results to the wider community. Therefore a loosely coupled architecture is essential to allow the flexibility of adding extra services at a later stage when the user requirements change.

A clear priority requirement in the project is the design of a user-oriented portal, aimed at the non-computer specialist, with accessible resources, i.e. scientific data and publications, that are easy to browse, upload and download. In view of this, portal technology is incorporated into the design of the CORE. A portal is described as "*a network service that brings together content from diverse distributed resources using technologies such as cross searching, harvesting, and alerting, and collate this into an amalgamated form for presentation to the user*" [12]. A portal can provide personalisation to users and allow them to access the VRE using Web browsers. Such features make the VRE easy to use for users who may not be computer literate.

Another important finding from the user requirement study is that different types of users require data to be presented differently. For instance, biologists are generally interested in graphical representations of data while mathematicians might prefer numerical interpretation of the same set of data.

The authors learned from the study that the users require infrastructure to run simulations and analyse large scale trial data. Hence, it makes sense to include Grid services in the CORE since the services can provide secure and managed access to distributed computational power.

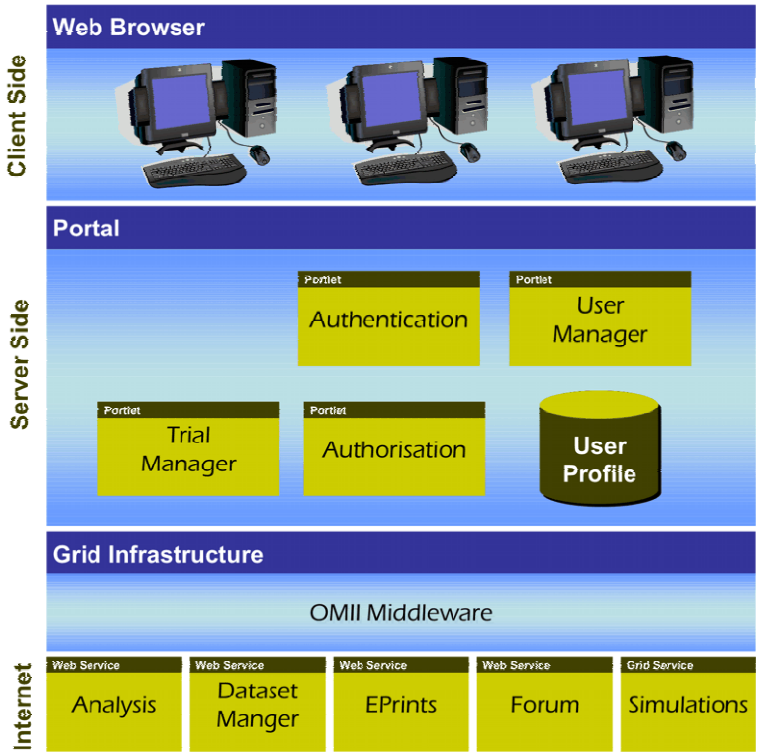

Figure 2: Overview of the CORE infrastructure

3.2 The CORE Infrastructure

The CORE is being implemented as a toolkit of generic components. Figure 2 illustrates the CORE framework concept, in the context of an e-science community Web site and integrated Grid/Web-based services.

OMII is a collection of tested, documented and integrated software components that provides a standard platform for integrating e-Science middleware. It also acts as a simple, secure web service-based Grid infrastructure for e-Scientists [11]. Hence, by using the OMII middleware, end users of the CORE toolkit can access Grid resources and applications in a trusted and secure environment.

Four Web services are factored based on the functionalities provided by the VOEU toolkit. The authors first decomposed the toolkit into processes and then identified which ones were potentially to be reused in other VREs. These reusable processes are *Analysis*, *Data Set Manager*, *EPrints* and *Forum*, which are the Web services depicted in Figure 2. These services will enable users to perform tasks such as formalising trial protocols, storing and analysing data, submitting and reviewing articles, and discussing research findings in a forum. Another Grid service, named *Grid Simulation* is also included in the VRE infrastructure. It will provide users with functions such as job submission, file transfer, and credential management in running their simulations.

A portal is used to facilitate the sharing of research resources in the CORE VRE. The main purpose of a portal in this infrastructure is to act as a presentation layer which aggregates, integrates, personalises, and presents information, transactions, and applications to the user according to their role and preferences. It provides a persistent state for an individual user or a group of collaborators. In the context of personalisation and embedding, portals achieve this through creating distinct pluggable components of functionality and offering them to the users as visible components. Each of the components can be constructed by utilising the emerging technology of Portlets. JSR-168 is the Java Portlet Specification adopted by the majority of Portal vendors in the market [1]. The specification aims to standardise the pluggable portal components so that they are independent of the portal vendor solution. Hence, the authors will design and develop the components of the of CORE portal based on the JSR-168 Specification.

4. Conclusions and Future Work

The design of the VRE in the CORE project described in this paper is underpinned by

Grid/Web services technology based on the concepts of SOA. The SOA theme is to enable reuse via shared services, where flexible granular functional components expose service behaviours accessible to other applications via loosely coupled standards-based interfaces. There is another common ways of integrating systems, which is to integrate at the user interface level using portals. However, the SOA approach does not preclude using portals, and is in fact agnostic about how the rest of the applications in an organisation are configured. Thus, SOA is a good approach for constructing the VRE.

The CORE VRE supports sharing and dissemination of findings, i.e. scientific data and publications, in research activities. This is important as any scientific effort will have diachronic features, i.e. collaborative research activities extending through time, enabling the influence of these activities to carry on beyond their timescale and enabling the dissemination of research findings beyond the original boundaries of collaboration. The CORE project should have a major impact in a number of areas including:

- Being able to keep track of the research administration: trial protocol, ethical approval, and workflow as the trial progress,
- Enabling access to research data from various trials and in formats that allow analysis of the data,
- Allowing easier meta-analysis or thematic reviews,
- Monitoring the effectiveness of surgical interventions,
- Enabling a consortium to write appropriate documents for dissemination (medical reports, journal articles, etc)
- Producing up to date learning and teaching material.

The CORE project will build a VRE demonstrator based on the CORE infrastructure presented in this paper. Issues such as semantic Grid/Web services will also be investigated in order to enhance the functionality of the proposed VRE, particularly in the areas of automation and organisation. For example, it maybe possible to automate the feasibility check of possible trials by looking at the availability of resources, such as laboratory timetables, datasets, and staff. In addition, evaluation of the demonstrator will be conducted qualitatively (using workshops and focus groups) and quantitatively (using experiments) once it is implemented. The evaluation will focus on the usability aspects of the demonstrator.

References

[1] R. Allan, C. Awre, M. Baker, and A. Fish (2003) Portals and Portlets 2003. National e-Science Centre, article available from: http://www.nesc.ac.uk/technical_papers/UKeS-2004-06.pdf

[2] Booth, D., Champion, M., Ferris, C., McCabe, F., Newcomer, E. and Orchard, D. (2004) Web Services Architecture (W3C Working Group Note 11 February 2004). World Wide Web Consortium, article available from: http://www.w3.org/TR/ws-arch/

[3] Buetow, K. (2005) Cyberinfrastructure: Empowering a "Third Way" in Biomedical Research, *Science*, 308(5723), pp. 821-824.

[4] Carr, L., Miles-Board, T., Wills, G., Power, G., Bailey, C., Hall, W. and Grange, S. (2004) Evolving a Digital Library Environment to the Changing Needs of its Users. In *Proceedings of Healthcare Digital Libraries Workshop 2004 (HDL 2004)*, Bath, UK.

[5] CORE Project Proposal, University of Southampton, article available from: http://www.core.ecs.soton.ac.uk/overview/

[6] Grange, S., Wills, G., Power, G., Miles Board, T., Carr, L. and Hall, W. (2003) Building a dynamic review journal (DRJ) - Extending the role of the Virtual Orthopaedic University. In *Proceedings of the 3rd Annual Meeting of the International Society for Computer Assisted Surgery (CAOS International)*, Marbella, Spain, pp. 122-123.

[7] Hey, T. and Trefethen, A. (2005) Cyberinfrastructure for e-Science, *Science*, 308(5723), pp. 817-821.

[8] Hitchcock, S., Carr, L. and Hall, W. (1996) A Survey of STM Online Journals 199095: the Calm before the Storm. In Mogge, D. (ed.) *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, 6th edition, Association of Research Libraries, Washington , Washington, D.C., pp. 7-32.

[9] The LHC Computing Grid, Web site: http://lcg.web.cern.ch/LCG/

[10] OASIS SOA Reference Model TC, OASIS, Web site: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=soa-rm

[11] OMII Web site: http://www.omii.ac.uk/

[12] Portals: Frequently Asked Questions. JISC, article available from: http://www.jisc.ac.uk/index.cfm?name=ie_portalsfaq

[13] Sim, Y.W., Wang, C., Gilberts, L. and Wills, G.B. (2005) User Requirement Study for a Virtual Research Environment. Technical Report, University of Southampton, article available from: http://eprints.ecs.soton.ac.uk/10969/

[14] Virtual Orthopaedic University, University of Southampton, Web site: http://www.voeu.ecs.soton.ac.uk/

[15] Wilson, S., Blinco, K. and Rehak, D. (2004) Service-Oriented Frameworks: Modelling the infrastructure for the next generation of e-Learning Systems. Joint Information Systems Committee (JISC), article available from: http://www.jisc.ac.uk/uploaded_documents/AltilabServiceOrientedFrameworks.pdf