



Report of the e-Frameworks Meets e-Science Workshop

Editors: Jennifer M. Schopf, NeSC and Argonne National Laboratory
Iain Coleman, NeSC

Contributing Authors: Matthew Dovey, Rob Allan, Gary Willis, Tony Linde,
Stephen Pickles, Alex Hardisty

Date: 28-4-2006

Abstract:

This document reports on the e-Frameworks Meets e-Science Workshop, held at the e-Science Institute on 28 April 2006.

The e-Framework is an international initiative from the U.K's Joint Information Systems Committee (JISC), Australia's Department of Education, Science and Training (DEST) and partners (<http://www.e-framework.org>). It aims to create a framework identifying the definition of services and procedures for the creation and re-use of software components and tools in a Service Oriented Architecture. Services may be common across the domains of e-Learning, e-Research (e-Science) and Digital Information. They may also be domain-specific.

This workshop addressed specifically the needs of the e-Science community, and the mapping of the services and use cases in that environment to the larger framework being defined. It consisted of an opening presentation, and then a day of discussion covering the various topic areas defined in the e-Frameworks documents.

Reference: NeSC report UKeS-2006-03 available from
http://www.nesc.ac.uk/technical_papers/UKeS-2006-03.pdf

Executive Summary

The e-Frameworks Meets e-Science Workshop was held on 28 April 2006 at the e-Science Institute.

The meeting addressed the following topics:

- What is the glossary being used in the e-Framework approach, and how can this be used in the e-Science environment?
- What is the scope of the e-Framework approach?
- What goal and roles do the e-Framework group envision for themselves in an e-Science environment?

A set of recommendations was elicited from the meeting, and given to the JISC e-Framework group. These recommendations included basic clarification of terminology and ways to present the approach, and the need for further discussion with domain experts on several points.

The following avenues for future work were identified:

1. A small set of terminology needs to be agreed upon.
2. A glossary is required, in part to help translate between e-Science, e-Learning, and others.
3. The background document needs to be clarified with respect to the goal of delivering interoperability. The outcome of the e-Framework Working Group is meant to be a tool to inform people: it is not a methodology to define interoperable standards and identify gaps. This needs to be clarified in the document.
4. Support services, production services and training should be added.
5. e-Science/e-Research should just become "Research".
6. There is concern about whether different terms need to be listed for multiple domains.
 - a. This needs further discussion with domain experts.
 - b. Presentational/sociopolitical aspects need to be considered as well as technical aspects.
7. The eFW group should look at GGF use cases and OGSA work.
8. A re-evaluation of the goals may be needed. While most people thought eFW might be appropriate, this proved to be unclear.
9. One approach is to proceed in two phases, with phase one dealing with classification, and phase two concentrating on community relevant issues.

Purpose of the Meeting

The e-Framework is an international initiative from the U.K's Joint Information Systems Committee (JISC), Australia's Department of Education, Science and Training (DEST) and partners (<http://www.e-framework.org>). It aims to create a framework identifying the definition of services and procedures for the creation and re-use of software components and tools in a Service Oriented Architecture. Services may be common across the domains of e-Learning, e-Research (e-Science) and Digital Information. They may also be domain-specific.

This workshop specifically addressed the needs of the e-Science community, and mapped the services and use cases in that environment to the larger framework being defined. It consisted of an opening presentation, and then a day of discussion covering the various topic areas defined in the e-Frameworks documents. The meeting took place on April 28, 2006 at the e-Science Institute in Edinburgh, UK.

The following topics were addressed at the meeting:

- What is the glossary being used in the e-Framework approach, and how can this be used in the e-Science environment?
- What is the scope of the e-Framework approach?
- What goal and roles do the e-Framework group envision for themselves in an e-Science environment?

Section 1 summarises the overview discussion that defined the e-Framework vocabulary and general approach. Section 2 provides a brief overview of the project summaries that were presented by the e-Science attendees. Following from these, Section 3 provides a summary of the ensuing discussion. Section 4 summarizes the meeting and its conclusions overall.

Additional information on the workshop, including the agenda, participant list, and talk slides, can be found online at <http://www.nesc.ac.uk/esi/events/676/>

1 Overview of e-Frameworks – Matthew Dovey

1.1 e-Framework Overview

The e-Framework is a set of concepts, methodologies and best practice guides for facilitating a service-oriented approach to technical interoperability within and across research, teaching, libraries and administration in higher education. The e-Framework has been established to help the education and research worlds to take advantage of the opportunities offered by a standards-based service oriented approach. The programme was initiated by JISC and Australia's Department for Education Science and Training (DEST) and is now attracting growing International support. Most recently New Zealand's Ministry of Education, and SURF in the Netherlands, have agreed to become involved.

The primary goal of the e-Framework is to facilitate technical interoperability within and across education and research through improved strategic planning and implementation processes.

A set of principles guides the development of the e-Framework:

- A service oriented approach to system and process integration
- Development, promotion and adoption of Open Standards
- Community involvement in development of the e-Framework
- Open collaborative development activities
- Flexible and incremental deployment

The context in which the e-Framework operates is the technical infrastructure that supports the communities active in education and research. There is a need for greater coherence in development, and for a map of what has been developed and the standards and specifications that underpin them. This information would enable a strategic approach to planning programmes of development, and would provide institutions with information on what is available and ready for adoption and mainstream use.

The e-Framework is intended to provide a strategic approach to technical infrastructure development within and across domains, as well as a consistent technical vocabulary. It is further aimed at providing a focal point for interaction with software developers and those providing services to education and research, and acting as a catalyst for the development of further specifications and standards

The benefits of the e-Framework can be seen from several perspectives. For e-Framework Partners it will provide a map of a complex environment, and a strategic planning tool for prioritised investment in both standards development and interoperability technologies. It should also result in an improved return on investment through coordination and collaboration between Partners.

For Institutions it will enable the alignment of strategies and infrastructure development to support the areas of education and research. It should give rise to a greater choice of systems and suppliers, an improved return on investment in existing systems, and more effective communications between communities through shared understanding. Interoperability within and across institutions and national boundaries will also be enhanced.

For Developers it will lead to a better understanding between suppliers and customers, more rapid development cycles through reusable components, and the entry of small innovative players into the market. The e-Framework will also result in a faster response to customer requirements, improved communication and collaboration between developers, and more flexible business models for software development.

The e-framework is documented by guides, methodologies and analysis. These include service definitions and descriptions and reference models.

A service oriented approach describes a way of building systems using discrete service components. This highly flexible approach reduces the costs of system development and maintenance through reuse of components. Reusing services in this way also improves interoperation between systems and simplifies the extension of system capability.

Services perform specific tasks. These can be made available over the Web – a well-known example of a service is Google Search. Others may be less well-known but could be just as useful – an example could be an authentication service (checking that you are who you say you are).

Reference Models are developed by communities to identify their particular learning, teaching or research requirements. Reference Models provide an analysis of the tasks involved, identify the sets of services that will be called upon, identify how they will work together, and provide all the information that developers will require to create the interoperability points for a particular applications.

In order for a range of services to be able to work together they must conform to a set of technical standards. Standards enable services and systems to ‘talk to each other’ and exchange data. At a simplistic level this is analogous to agreeing to country codes that enable international telephone communications.

The following are within scope of the e-framework:

- The methodologies for describing and presenting a reference model
- The registry of reference models
- The identification and documentation of service patterns - service patterns will be derived from the study of multiple reference models
- Infrastructure to support community contributions

- Quality Assurance
- Development and presentation of a consistent view

The e-framework is currently concentrating on identifying service genres – high level service types – within the following areas: learning and teaching services, research services, library services, IT services, administrative services, and common services across domains (e.g. security).

It is expected that community activities (not the e-framework) will work at less abstract levels below the service genres.

References

"The e-Framework for Education and Research: an Overview", Bill Olivier, Tish Roberts, Kerry Blinco, July 2005
<http://www.e-framework.org/resources/e-FrameworkrV1.pdf>

1.2 e-Framework for Research

Jane Hunter (DEST) and Matthew Dovey (Oxford University, now JISC) have attempted to map the research specific service genres, based on previous work e.g. "Service Classes for the e-FER", Rob Allan CCLRC e-Science Centre (for JCSR)
<http://www.Grids.ac.uk/Papers/Classes/classes.html>

The e-Learning Framework - Common Services
http://www.elframework.org/common_services

The e-Learning Framework - Learning Domain Services
http://www.elframework.org/learning_domain_services

DEST, JISC-CETIS and Industry Canada "Service-oriented Frameworks: Modelling the Infrastructure for the next generation of eLearning Systems"
http://www.jisc.ac.uk/uploaded_documents/AltilabServiceOrientedFrameworks.pdf

The work had identified an initial set of genres which could be deemed specific to research and IT services:

e-Science/e-Research Service Genres

- Remote Instrument
- VR/Immersive
- Visualisation
- e-Lab notebook
- Simulation/Modelling
- Computational

Collaboration Service Genres

- AV Conferencing
- Access Grid
- Whiteboard
- Chat
- Email
- Presence
- Forum
- Messaging
- Wiki
- Application Sharing

Organisational Service Genres

- Project Management
- Personal Organisation
- Personal Development

Grid/IT Service Genres

- System Management
- Service Management
- Software Management
- User Management
- Group/VO Management
- Identifier/Resolver
- Network Management
- Job Management
- Resource Management

There are other genres important to research but these are shared with other areas of the e-framework so were represented in this context. A strong message from the meeting is that this presentation suggests that these areas are “owned” by other areas, which is not the intention, and a better way of representing these genres as being important within research whilst recognising overlap and commonality at the relatively high level of abstraction is retained.

2 e-Science Use Cases

2.1 Rob Allan, e-HTPX

The e-HTPX project (www.e-htpx.ac.uk) is developing an architecture and middleware services deployed as a “gateway for experimental facilities”, which in the future will include Diamond. The workflow of the project comprises six stages, as follows:

1. protein production – expression of gene sequence as requested by users
2. crystallisation – protein in solution in 96-well sample trays allowed to crystallise under various conditions, which may take days or months. A service is deployed which allows remote visual checking of crystals using a robot and camera.
3. data collection – Prior to data collection the crystals are labelled using a bar code and categorised. A shipping service is used to send the samples in a liquid-nitrogen filled Dewar and track them. They must then be maintained prior to scheduling an experiment.
4. phasing - Data collection per se consists of exposing the samples to X-ray light under strictly controlled conditions and collecting diffraction images (intensities) and corresponding meta-data. An automatic process is available as a service to do this and also controls image quality so that only as much data is collected as is required for subsequent analysis. Once intensities are available, phasing can begin.
5. protein structure determination – intensities plus phases results in a set of data which can be converted into a 3D structure.
6. data deposition – upload of 3D structure information and meta-data to the Protein Data Bank at EBI, Hinxton. Available for international researchers to use.

Real artefacts, such as solution well trays, crystallised protein samples, cameras, robots and X-ray beam facilities are involved in this project, in addition to data management, computational analysis, metadata collection and database uploading.

2.2 Gary Wills, CORE

The collaborative orthopaedic research environment (CORE, www.core.ecs.soton.ac.uk) is a JISC funded project, under the Virtual Research Environments programme. The project aims to provide integrated computer support as a part of the requirements of the surgeon's Continuing Professional Development, across the research and educational cycles, because these activities are intrinsically coupled. The portal allows surgeons to create technical material (including non research material for education), analyse data (from their own trials or hospital data), investigate hypotheses (from their own work or as meta or thematic reviews), discuss the findings from their own or others' work, and prepare and submit articles for review.

The CORE project uses a Service Oriented Architecture and a portal framework to provide the user interaction. The system is currently in use in two hospitals. As some of the analysis from the laboratory based orthopaedic departments is computationally intensive, the system will also connect to the Grid via the portal framework. It is our intention to use the OMII middleware to enable us to do this.

2.3 Tony Linde, AstroGrid

AstroGrid (www.astroGrid.org) is a PPARC-funded e-Science project aimed at providing a Virtual Observatory (VObs) infrastructure which makes worldwide astronomical data and software easily available to all astronomers via standards-based

components. There are several areas in the VObs field where “e-Frameworks meets e-Science”.

Although most astronomical data is freely available, much of the most interesting data is proprietary to principal investigators and their collaborators. This requires that we are able to authenticate users and authorise their use of data and services. The international VObs community of projects (IVOA: <http://ivoa.net>) is defining a way to do this without individual certificates, probably based around MyProxy and associated technologies. Ultimately, however, users will expect that their institutional single sign-on facility will also give them access to VObs and Grid services, so we will have to think about integrating VObs security with Shibboleth and e-Frameworks services.

Another service that AstroGrid has developed and that is being standardized by the IVOA is that of virtual storage. This is a means by which users can have personal access to data storage facilities without regard to the location of these facilities. At the moment the IVOA is defining the standards for using a facility like this such that that the storage could be as simple as a filestore on a server or as complex as a distributed storage facility like SRB. This might be a facility that could become part of e-Frameworks.

Finally, the AstroGrid services are made available through a client-based Astronomer's Workbench. This launches several own-built tools and provides access to 3rd party tools, all made interoperable using an AstroGrid-developed common interface (PLASTIC: <http://sourceforge.net/projects/plastic>). It would be our long term aim to either develop the workbench into a VRE or integrate workbench facilities into an existing VRE, thus providing access for UK astronomers to institutional and research facilities such as library and resource browsing from the same platform.

2.4 Stephen Pickles, Use Cases

National Grid Service

The NGS provides compute and data resources to UK users and user communities. Its security model is based on a Public Key Infrastructure, with extensions (GSI) to support delegation. In contrast, the dominant security technology in the broader JISC domain has been Athens, which is steadily giving way to Shibboleth (and PERMIS). Viewed from a sufficiently high level, these technologies achieve the same goals – single sign-on, authentication, and authorisation. However, the domain of single sign-on in the two universes is different: GSI emphasises programmatic access to resources and services, whereas Athens and Shibboleth, being designed with Web browsers as the main delivery vehicle, are difficult to re-use programmatically.

To make broader JISC-funded resources (such as centrally curated datasets) accessible in a Grid context, the "services" and capabilities must become programmatically accessible (e.g. by becoming Web services in the technical sense) while preserving the original access control restrictions (which implies that enforcement must now be performed at the service layer). Attempting to bridge the AA universes in this direction gives rise to various issues, such as making Shibboleth

attributes accessible from within a Grid or Web service, a problem being addressed by the US GridShib project. Specific examples are cited below.

To increase the relevance of NGS to a broader community, through e.g. provisioning of domain-specific portals using NGS resources, one must build bridges from the AA universe of Shibboleth to the GSI universe of NGS. This is the "ShibGrid" problem, under investigation in two JISC-funded projects, ShibGrid and SHEBANGS.

Grid resource providers also face issues such as accounting, logging, configuration management, booking/reservation, user management. Such terms also exist within the nomenclature of e-Frameworks. However it is not clear whether the coincidence of terminology is due merely to the analogous operations that they describe, or whether there is a real potential for re-use of software or services across an e-Framework enlarged to include Grid.

Grid access to social science data sets

A series of e-Social Science pilot demonstrator projects, ConvertGrid and GEMEDA (now complete), and most recently, the JISC-funded GEMS project (Grid Enabling MIMAS Services) share the common theme of making centrally curated datasets accessible programmatically within a Grid environment. It was noted that:

- Users must register with data owners to access many strategically important datasets such as census, satellite data, and the like.
- This has resulted in a wealth of access control information being held in Athens-based systems, and these are slowly migrating to Shibboleth.
- This access control information is not readily accessible from a Grid context (this is the GridShib problem).
- Access is currently enforced by data portals. This must shift to the service layer to make the data Grid-accessible.

Nektar, SPICE & Vortronics (2005)

Nektar, SPICE & Vortronics were three projects jointly funded by the EPSRC and the US NSF for application exemplars exploiting the federated resources of the UK NGS and the US TeraGrid. All involved parallel codes and needed high-bandwidth cross-site communications. Nektar and Vortronics used MPICH-G2, and Grid-enabled flavour of MPI that permits a single parallel application to run distributed across remote supercomputers and/or clusters. SPICE used computational steering software from the EPSRC-funded RealityGrid project. All three require simultaneous access to multiple computational and network resources, and in the case of SPICE, visualisation resources and sometimes AccessGrid nodes. The lack of convenient facilities for advance reservation and co-allocation is a significant pain point for these projects, and is arguably a major obstacle that is retarding some of the more imaginative uses of Grid computing.

NanoCMOS-Grid

NanoCMOS-Grid is a third-round EPSRC-funded e-Science pilot project, due to start in October 2006. It will generate a significant cottage industry in turning chip design tools and design flows into Web services and workflows. Issues it will face include:

- secure invocation of services (authorisation must protect access to software licenses and to the underlying computational resources)
- provenance
- markup/metadata/discovery
- security of data (sensitive IP is involved and it is not acceptable to trust systems administrators).

2.5 Alex Hardisty, RAVE

RAVE (Resource-Aware Visualisation Environment, www.wesc.ac.uk/projectsite/rave) is a wide-area distributed collaborative visualisation service. It permits a user to publish and announce a dataset for visualisation. Other users may discover the dataset and join a collaborative visualisation session based around that dataset. It is not necessary for subscribers to have local rendering capabilities at their physical location. Users without such rendering capabilities may discover an appropriate rendering service provided by RAVE and may nominate that service to render the published data for them.

The actors in this scenario are the owner of the dataset (local user) and multiple remote users wishing to join the collaborative visualisation session.

The functional components that make up the RAVE service are:

- The “Yellow Pages” service, which discovers the location of RAVE services and enables the selection of the most suitable service or resource.
- The Data Service, which hosts and distributes data, synchronizes multiple collaborating clients, and acts as a bridge between RAVE and 3rd party applications.
- The Render Service, which performs rendering for under-resourced clients.
- The Active Client, which performs local rendering and interaction with data, and adapts to local hosts’ resources.
- The Thin Client, which performs local display and interaction via remote rendering, and adapts to prevailing network conditions.

3 Recommendations of the Meeting

Several possible avenues for future work were identified as a result of the meeting.

1. A small set of terminology needs to be agreed upon.
2. A glossary is required, in part to help translate between e-Science, e-Learning, and others.
3. The background document needs to be clarified with respect to the goal of delivering interoperability. The outcome of the e-Framework Working Group is meant to be a tool to inform people: it is not a methodology to define interoperable standards and identify gaps. This needs to be clarified in the document.
4. Support services, production services and training should be added.
5. e-Science/e-Research should just become "Research".
6. There is concern about whether different terms need to be listed for multiple domains.
 - a. This needs further discussion with domain experts.
 - b. Presentational/sociopolitical aspects need to be considered as well as technical aspects.
7. The eFW group should look at GGF use cases and OGSA work.
8. A re-evaluation and clarification of the goals may be needed. While most people thought eFW might be appropriate, this proved to be unclear.
9. One approach is to proceed in two phases, with phase one dealing with classification, and phase two concentrating on community relevant issues.

4 Summary and Conclusions

The e-Framework Working Group aims to create a framework identifying the definition of services and procedures for the creation and re-use of software components and tools in a Service Oriented Architecture. This workshop examined how the approach might be used in an e-Science domain, and evoked much discussion on how this work could be presented to a larger set of e-Scientists for additional feedback.

A set of recommendations was elicited from the meeting, and given to the JISC e-Framework group. These recommendations included basic clarification of terminology and ways to present the approach, and the need for further discussion with domain experts on several points.

Acknowledgements

This work was supported in part by the Joint Information Systems Committee (JISC). We also acknowledge the support of the National e-Science Centre in the UK for their assistance with the local arrangement for this workshop.