

Resource Discovery Services for Grid Computing Training

Boon Low¹ David Fergusson¹ John MacColl²

¹*Training, Outreach and Education Division, National e-Science Centre,
15 South College Street, Edinburgh EH8 9AA, UK
boon.low@ed.ac.uk, dfmac@nesc.ac.uk*

²*Digital Library Division, Edinburgh University Library, University of
Edinburgh, George Square, Edinburgh EH8 9LJ, UK
john.maccoll@ed.ac.uk*

Abstract

Resource discovery is a core function of e-learning. While initially confined to library environments, it has become pervasive, partly due to the advent of web-based searching and also service-oriented approach. The latter is the focus of this paper as resource discovery (web) services enable existing resources to be identified from a source (library) and re-purposed for educational purposes in other systems such as virtual learning environments. This paper describes the development of such resource discovery services developed as part of an initiative to pilot e-learning and a shared digital library infrastructure for Grid Computing training projects in Europe. The development is also related to a project funded by the UK Joint Information Systems Committee (JISC) on publisher metadata interoperability. The resource discovery services correspond to an emerging international service-oriented framework for developing e-learning. This paper also provides an overview of the framework and the use-scenarios in which the resource discovery services have been deployed.

Keywords: Resource Discovery Services, e-Learning Framework, Digital Library, Grid Computing Training

1. Introduction

A trend exists in which service-oriented and distributed computing technologies are increasingly being adopted for developing e-learning infrastructures. Generally, service-oriented approaches involve the development and use of web services: both new and derived from legacy systems. Once web services are exposed through common system interfaces, they can be consumed by custom-built applications to facilitate a variety of educational processes, some of which may previously have been limited by closed- and organisational system boundaries. For example, library metadata can be repurposed as reading lists in virtual learning environments (VLEs), through applications that are capable of cross-searching the library catalogues by means of resource discovery web services.

The E-Learning Framework (ELF) is an international initiative to build a common approach towards service-orientation in e-learning [1][2]. It is not an initiative to provide an overarching solution with detailed system architecture *per se*. The main focus is the identification and the factoring of e-learning web services. ELF identifies two broad sets of services:

- *Learning Domains Services* - specific to e-learning such as assessment, learning activities design, learner and course management
- *Common Services* - core services such as resource discovery and authentication services to underpin the learning domain services; these are shareable for e-learning as well as other purposes, e.g. digital library and e-research.

Services from both strands may correspond to existing or emerging technical standards developed by standards bodies such as the ISO, the IEEE and the IMS Global Learning Consortium [3]. Adherence to technical standards is essential to ensure interoperability and the ease of service consumption among disparate systems. ELF therefore also has a role in encouraging standards development and adoption, in particular piloting the use of emerging data schemas and web service interfaces.

This paper describes a shared digital library infrastructure that is based on a subset of ELF Common Services. The main focus is on the development of the resource discovery services (part of the infrastructure) in collaboration with a project funded by the UK Joint Information Systems Committee (JISC) [4]. The JISC project has developed two ELF Common Services - *search* and *resolver* services, advancing the development of two resource discovery service standards while investigating how the services can be utilised to enable metadata aggregation and enhancement both remotely and locally. The digital library infrastructure also underpins e-learning pilot implementations for two Grid Computing training projects in Europe [5][6], facilitating shared services for various portals and virtual learning environments.

2. A Digital Library for Grid Computing Training in Europe

Grid Computing is currently one of the most prominent areas of research and development. As an advanced form of network infrastructure, underpinning a diverse range of e-Science applications, grid infrastructure enables the sharing of distributed computing resources and data in virtual networks – the so called “Virtual Organisations”. Grid infrastructure typically provides “higher throughput computing by taking advantage of many networked computers to model a virtual computer architecture that is able to distribute process execution across a parallel infrastructure”[7].

Enabling Grids for E-science (EGEE), a project funded by the European Commission's Sixth Framework Programme, is a leading initiative in grid computing involving 70 organisations from 27 countries [5]. One of the project objectives is to engage a wide range of users from science and industry by providing them with extensive technical and training support. The provision of training is therefore a core activity of the EGEE project. Currently, training activities are delivered by a federated training team, consisting of experts from 22 partner organisations who organise events at distributed locations across Europe and in other countries. One of the challenges for the EGEE training team is to provide training that targets the needs and skills of a diverse community - from new staff on the project, to academics, researchers and new users from the industry who wish to exploit the potential of grid computing. Another challenge is to make the training more widely accessible beyond the confines of training events that are bound to specific times and various geographical locations. As a result, the EGEE training team is undertaking a pilot implementation to harness the benefits of e-learning, insofar as using ELF to provide a basis for developing pilot services. The first phase of the pilot resulted in the development of a service-oriented digital library infrastructure mentioned in this paper.

3. Resources and Discovery Services

While some e-learning resources may be readily accessible on the web, most reside in digital repositories, VLEs and dedicated portals, beyond the reach of web search engines. The following is an overview of the type of resources that are relevant to grid computing training purposes:

- Training materials (‘learning object’) such as code exemplars, exercises, manuals, tutorials, exam papers, lectures and presentations in various formats e.g. PDF, PPT, audio/video files
- Bibliographical reference objects such as books, journals, articles, websites

The training materials can be created in a federated manner, for example by a distributed training team in the EGEE project. These are available with basic metadata and stored in distributed repositories. Bibliographical reference objects are becoming available, due to the publications of project results and dissemination. As Grid Computing technology advances and becomes embedded in formal university education [6], the scope of training is extended to the teaching of concepts and fundamental computing skills. Hence, there is a requirement to expand the digital library with bibliographic and reference objects. This strand of development is being accomplished collaboratively with the JISC *metadata+* project funded under the Publishers and Library/Learning Solutions (PALS) Metadata and Interoperability Programme [4]. The project curates metadata related to informatics subjects for discovery purposes and enhances the metadata from remote data sources and user annotations.

Given the widespread use of Google, federated-searching (across multiple data sources simultaneously) is fast becoming the most popular approach to resource discovery. Unifying accesses to heterogeneous repositories is generally hindered by the fragmentation of the information environment that is typified by widely-distributed and autonomously-maintained services. These are in effect ‘information islands’, unconnected to an available common source - examples include the list of

databases in a digital library [8]. Efforts to bridge these information islands remain a challenge due to the heterogeneity of repositories and access methods, where these exist.

The digital library needs to be based on a repository that provides a unified resource discovery mechanism matching users and their intended resources in a direct and rapid manner [9]. This necessitates the development of a *union* metadata and content repository, consolidating the diverse material sources related to the EGEE project. The resource discovery services correspond to the following ELF common services:

- Search service - supports the finding of e-learning resources stored at the digital library through the URL- and SOAP-based web services - see “Technical and Metadata Development”.
- Resolver service - provides services based on the use of the OpenURL Framework [10], for direct linking (from other websites) to the resource using context-sensitive metadata such as ‘author’ and ‘title’ fields.

4. Technical and Metadata Development Experience

The digital library infrastructure is based on an open source and service-oriented repository system – Fedora [11]. The system provides a set of off-the-shelf web services - Fedora APIs (SOAP-based) for content interrogation (search/access) and management. Additional software engineering efforts were required to ensure that the repository system meets the main requirements of the digital library. The developments include modifying the Fedora source code and web service interface (Fedora API WSDL) to enable the total hits number to be included in the search results and page navigation by means of a start record number within a results set range.

The Fedora API services are proprietary. The resource discovery interface familiar to the digital library community is the Search and Retrieve URL (SRU) and OpenURL – both are service protocols that enable search queries to be sent via the web (using URL calls) with search results returned in various XML metadata formats [10][12]. In order to enable Fedora to be searched by a wider community, an existing SRU search engine from a previous JISC-funded project was deployed to broker the SRU access mechanism for Fedora. This involved integrating the search engine with the Fedora service API such that the proprietary search operation are exposed through a standard SRU system interface façade [13].

Metadata for the grid computing training resources was procured from grid computing websites, EGEE project partners and publisher sources, through data harvesting and batch processing methods. Software scripts were developed to parse training event websites to ascertain the appropriate resource descriptions and the locations of distributed training materials. Publisher metadata was sourced from cross-searchable bibliographical sources including library catalogues and publisher websites (listings). The procured metadata, in various formats, was mapped to the Dublin Core (DC) metadata elements [14] which is the default schema intrinsic to the Fedora storage system. DC is also an ISO standard that is widely used in digital library communities.

5. Services Use Scenarios

The resource discovery services can be utilised to support a range of web-based scenarios as described below:

- Digital library and portals: The EGEE Library (Figure 1) is a standard web application providing a simple user interface to the ELF resource discovery services [15]. It consumes the web services and renders the XML search results using XSLT stylesheets. Other EGEE project partners and other grid computing projects are currently building different (localised) service consumers using the same underlying web services, for example the ICEAGE project repository [6]. The shared infrastructure enables portals to be developed in a rapid manner since most efforts are spent on developing user interfaces that consume the web services, instead of infrastructure building.

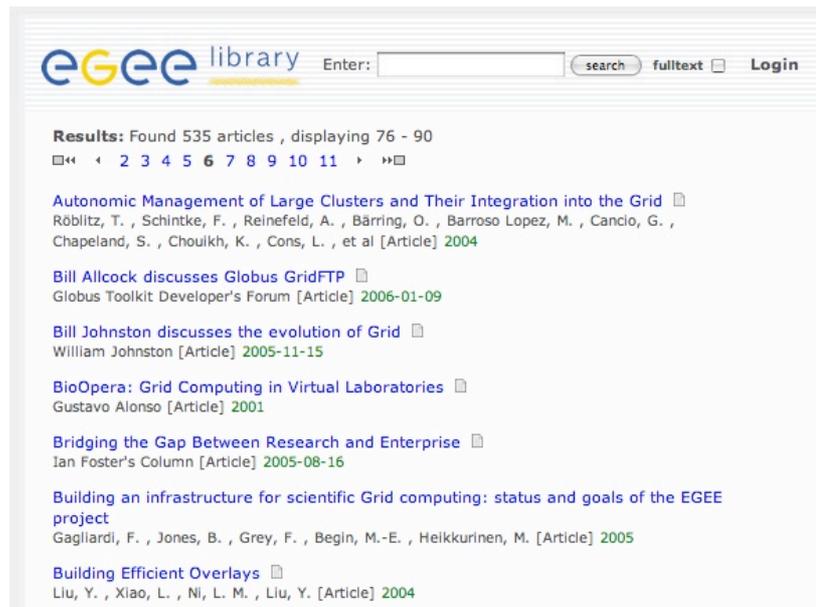


Figure 1. EGEE Library, a web client consuming the resource discovery services

- VLEs integration: The resource discovery services enable the library resources to be seamlessly cross-searched and repurposed within VLEs such as WebCT [13]. Figure 2 shows a VLE prototype currently being developed for the second phase of the EGEE e-learning pilot. It is based on the Gridsphere portlet container [16] and contains a digital library portlet consuming the same ELF resource discovery services.

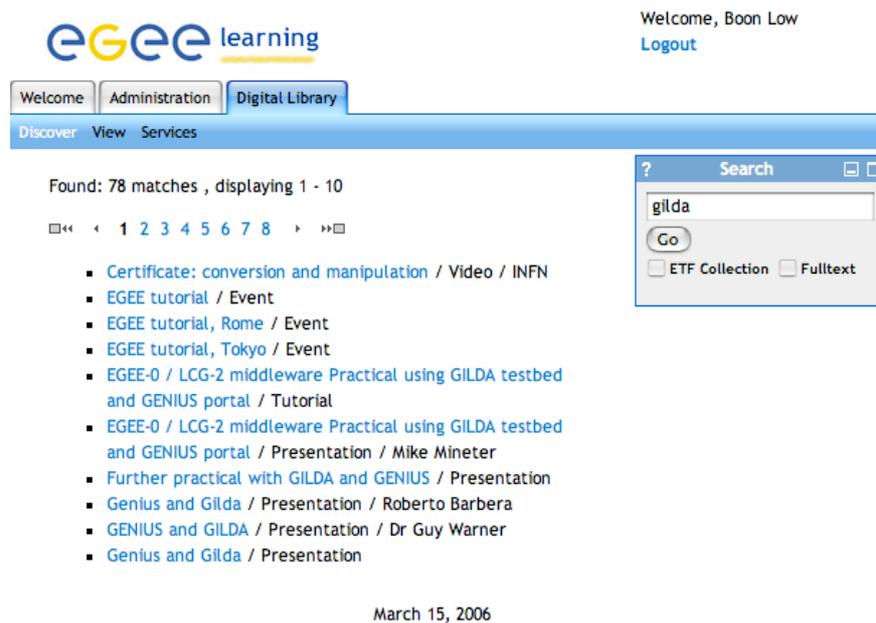


Figure 2. A portal consuming the ELF resource discovery services through a portlet

6. Conclusion

Resource discovery is an important function of e-learning. Recent work on open architectures and web services has allowed the initially closed virtual environments to be open and more interoperable. In this way, library services have been proactively developed in order to be flexibly retro-fitted to learning environments. It is therefore more conducive to search library resources from a variety of portals including virtual learning environments and have the results displayed and repurposed in a manner consistent with educational requirements. This paper described part of an initiative to advance

service-orientation in e-learning for Grid Computing projects in Europe. The outcome of the initiative has resulted in a common digital library infrastructure based on the e-Learning Framework (ELF) resource discovery services. While anticipating additional services development such as metadata annotation, this paper demonstrated briefly the efficacy of shared resource discovery services enabling different applications and portals to be developed in a variety of federated scenarios.

Acknowledgement

This work described in this paper was supported by the EGEE project under by the European Commission's Sixth Framework Programme (INFSO-RI-508833), and the *metadata+* project under the UK JISC Publishers and Library/Learning Solutions (PALS) Metadata and Interoperability Programme.

References

- [1]. S. Wilson, K. Blinco, D. Rehak, "Service-Oriented Frameworks: Modelling the Infrastructure for the Next Generation of e-Learning Systems", *Alt-I-Lab Conference*, 2004.
- [2]. <http://www.elframework.org/frameworks/general/frameworks/index.html>, A Technical Framework to Support e-Learning, 2004.
- [3]. <http://www.imsglobal.org>, IMS Global Learning Consortium.
- [4] http://www.jisc.ac.uk/index.cfm?name=project_metadata_plus, metadata+ project, "Machine Services for Metadata Discovery and Aggregation," JISC Publishers and Library/Learning Solutions (PALS) Metadata and Interoperability Programme, 2005.
- [5]. <http://www.eu-egee.org/>, Enabling Grid for e-Science project.
- [6]. <http://www.iceage-eu.org/>, International Collaboration to Extend and Advance Grid Education (ICEAGE) Project.
- [7]. http://en.wikipedia.org/wiki/Grid_computing, Grid Computing, Definition from wikipedia.
- [8]. L. Dempsey, R. Russell, "The Library, the catalogue, the broker: brokering access to information in the hybrid library", *The New Review of Information Networking*, Vol. 5, pp. 3-26. 1999.
- [9]. L. Dempsey, "The (Digital) Library Environment: Ten Years After", *ARIADNE*, Issue 46, February, 2006.
- [10]. <http://www.openurl.info/registry/>, Registry for the OpenURL Framework.
- [11]. C. Lagoze, S. Payette, E. Shin, C. Wilper, "Fedora: An Architecture for Complex Objects and their Relationships", *Journal of Digital Libraries, Special Issue on Complex Objects*, Vol. 6, No.2, pp. 124 – 138, 2005.
- [12]. <http://www.loc.gov/standards/sru/>, Search/Retrieve via URL (SRU).
- [13]. B. Low, J. MacColl, "Searching Heterogeneous e-Learning Resources", paper presented at the *DELOS Workshop, Digital Repositories: Interoperability and Common Services*, May, 2005. <http://delos-wp5.ukoln.ac.uk/dissemination/pdfs/low.pdf>
- [14]. <http://dublincore.org/documents/dces/>, Dublin Core Metadata Elements.
- [15]. <http://egee.lib.ed.ac.uk/>, EGEE Library.
- [16]. <http://www.gridisphere.org/gridisphere/wp-4/Documents/France/gridisphere.pdf>, J. Novotny, M. Russell, O. Wehrens, "GridSphere: An Advanced Portal Framework".