



Feedback from community on ILDG Metadata Catalogue Web Service Version 1

Project Title: QCDGrid2

Document Title: Feedback from community on ILDG Metadata Catalogue Web Service
Version 1

Document Identifier: D3.6_QCDGRID_MDWS_V1_REVIEW_1.0

Document Filename: D3.6_QCDGrid2-MDCWS-V1-Review.doc

Distribution Classification: Public

Authorship: Mark G. Beckett

Approval List: QCDgrid Development Team

Distribution List: UKQCD Collaboration

Document History:

<i>Personnel</i>	<i>Date</i>	<i>Summary</i>	<i>Version</i>
MGB	16 th May 2006	First release	1.0

Contents

1	Introduction	3
2	Development of ILDG MDC Web Service (Version 1)	4
3	Status of the ILDG MDC Web Service (Version 1).....	5
3.1	Status of the specification.....	5
3.2	Status of implementations	5
3.3	Feedback on the specification and implementation.....	5
4	Conclusions and further work.....	7
	References	8

1 Introduction

The QCDgrid project [8] is a core activity of UKQCD, a collaboration of UK academics and researchers that aims to procure and jointly exploit computing facilities for lattice field theory (commonly referred to as Lattice QCD) calculations. The primary aim is to increase the predictive power of the *Standard Model of elementary particle interactions* through numerical simulation of *Quantum Chromodynamics*. Such numerical simulations produce significant amounts of data and the purpose of the QCDgrid project is to provide software and supporting infrastructure that simplifies the management, storage, and manipulation of this data.

In the first three years of the project (2002 – 2004), software engineers at EPCC developed QCDgrid [10]—a data management system that combines the distributed resources of the collaborators into a robust facility called the *UKQCD Grid*. The result is a multi-terabyte storage facility over six UK sites at: Edinburgh (including the University of Edinburgh Advanced Computing Facility), Liverpool, RAL, Southampton, and Swansea. Glasgow is also a member of the consortium.

The facility is based on commodity hardware and open-source software. The hardware consists primarily of high specification, PC-based servers running the Linux operating system and managing large RAID storage arrays. On top of this infrastructure, the QCDgrid software (built using components from the Globus Toolkit, EGEE application stack, and an XML database) provides *Datagrid* management and user functionality – furnishing a simple and intuitive environment that hides the complexities of the underlying grid and presents a standard file system to the user. It incorporates a robustness metric that automatically disperses datasets across the grid, providing a resilience that ensures data is not affected by the loss of one (or possibly more) storage nodes. Security is leveraged from the Globus Toolkit, based on X.509 digital certificates issued by an approved Certificate Authority. The result is a reliable, secure data management system.

UKQCD is an important contributor to the International Lattice Data Grid (ILDG) [4], a group of like-minded scientists around the world who aim to share their data to accelerate scientific progress in the field of Lattice QCD. The ILDG was initiated in 2002 and, at the time of writing, has significant representation from research groups in Australia, France, Germany, Italy, Japan, UK and USA.

The ILDG infrastructure is being assembled as a web services layer that will aggregate the particular resources of contributing collaborations (for example, the UKQCD Grid) for the benefit of the wider community. To achieve its objectives, ILDG has established working groups to:

- Facilitate data sharing, through the standardisation of the form and content for Lattice QCD scientific data and associated metadata.
- Produce a set of specifications that define an architecture for an international *Grid of Grids* for Lattice QCD.

The contribution of the UKQCD to the ILDG is made, in part, by Work Packages 3 and 4 of the QCDgrid project [9]¹. These work packages define contributions to the development of a distributed Metadata Catalogue (MDC) and Replica Catalogue (RC), respectively, made through the ILDG Metadata Working Group [4] (the Working Group).

This report represents the main deliverable from Work Package 3.6 of the QCDgrid project [9]. Within this report, we appraise the status of the ILDG MDC Web Service Version 1 (the Service), as the Working Group members approach the realisation of a production service. The appraisal includes:

- A chronology of the key events in the development of the service.
- A status report on the situation, at the time of writing.
- A number of proposals for continued development and production deployment of the service across the ILDG community.

¹ Throughout this document, the terms “UKQCD contribution” and “QCDgrid project contribution” are both used to define the contribution made by the QCDgrid project to the ILDG.

2 Development of ILDG MDC Web Service (Version 1)

Below, we provide a brief chronology of the key events in the development of Version 1 of the web service (October 2005 – May 2006).

- **October 2005** – The formal specification for the ILDG MDC Web Service was initiated at the ILDG Middleware Working Group Meeting at the Centre for Computational Science in the University of Tsukuba in Japan on 27th and 28th October. This process built on high level discussions as to the form and function of the service that had taken place over the 18 months leading up to this meeting.
- **December 2005** – Based on the progress made at the Tsukuba meeting, UKQCD proposed a specification for the service [1] that was complemented by a WSDL² definition for the interface provided by the Japanese collaboration. The specification was provisionally accepted, subject to a small number of minor amendments to details, at a tele-conference in December 2005.
- **February 2006** – Using the specification and accompanying WSDL document as a guide, four of the ILDG collaborations – LQCD (Germany/Italy), JLDG (Japan), UKQCD (UK), and USQCD (USA) – began to design and develop a candidate implementation of the MDC Web Service interface on top of their existing infrastructures.

In the case of the QCDgrid project, the Web Service design [2] was released to the Working Group at the end of January 2006. Based on this design, an implementation was developed and deployed onto the QCDgrid Development Service during February 2006.

- **March 2006** – At the beginning of March, representatives from UKQCD (the QCDgrid project team) and LQCD (Dirk Pleiter from DESY Zeuten) met in Edinburgh to perform basic interoperability experiments on the candidate services deployed by the two collaborations. Based on the information gathered during these experiments, ambiguities and inconsistencies in the specification were eliminated and revisions to the MDC web service proposed, which received subsequent approval from other members of the Working Group.

The LQCD team also demonstrated two simple, web service clients that they had developed, both capable of interacting with a candidate web service implementation. These client tools greatly facilitated subsequent interoperability testing.

At the end of March, UKQCD proposed a formal test suite [3] for confirming the compliance of candidate implementations of the Web Service.

- **April 2006** – UKQCD released a revised version of their implementation of the web service. Following the formal test suite, it was confirmed that the UKQCD, LQCD, and USQCD implementations were consistent in terms of operation and functionality.
- **May 2006** – the Working Group presented an update on the ILDG MDC Web Service to the ILDG 08 [6] meeting. Based on the presentation, it was decided that the ILDG infrastructure should be presented at the Lattice 2006 conference.

Following on from ILDG 08, the Working Group formally released Version 1 of the specification.

² Web Services Description Language (WSDL) is the standard format for describing a web service. A WSDL definition describes how to access a web service and what operations it will perform. See <http://www.w3.org/TR/wsdl> for more information.

3 Status of the ILDG MDC Web Service (Version 1)

3.1 Status of the specification

The specification for the Service consists of three separate documents:

- A **WSDL definition** of the service interface [5] that must be exposed by an implementation.
- A **behavioural specification** for the service [7] that complements the WSDL definition, improving clarity and eliminating potential ambiguities.
- A **test suite definition** [3] that can be used to determine the compliance to the specification of candidate implementations of the service. This test suite is also supported by a script-based implementation that can be downloaded from the ILDG website [4].

Together these three documents represent Version 1.0 of the ILDG MDC Web Service specification.

3.2 Status of implementations

At the time of writing (May 2006), four collaborations – LQCD, JLDG, UKQCD and USQCD – have deployed implementations of the ILDG MDC Web Service Version 1. The UKQCD candidate has been deployed on the QCDgrid Development System, though is anticipated to be migrated to the production UKQCD Grid before June 2006. All four implementations comply with the test suite [3].

Two simple web clients have also been developed to demonstrate the functionality of the service. The LQCD service is pictured in Figure 1.

Looking to the future, the Lattice QCD group in Australia – who have an existing, though non-compliant web service for their MDC – aim to converge their implementation with the ILDG specification during the latter half of 2006. This will increase the number of collaborations running ILDG services to five.



Figure 1: Screenshot of the LQCD web client displaying ensemble metadata that has been retrieved from the MDC service hosted on the QCDgrid Development System.

3.3 Feedback on the specification and implementation

At this early phase, feedback regarding functionality has been gathered from a small number of expert users, both within and without the Working Group – specifically, Balint Joo (USQCD, Jefferson

Laboratories), Chris Maynard (UKQCD, University of Edinburgh), Dirk Pleiter (LQCD, DESY Zeuten), and Mitsuhsa Sato (JLDG, CCS Tsukuba). The feedback received from these users can be summarised, as follows:

1. The functionality defined in the specification supports a subset of the Query Use Case defined by the ILDG. Specifically, the current interface permits a user to retrieve either: a list of ensemble URIs / Logical Filenames (LFNs) from the ensemble / configuration repository; or a set of complete QCDML documents from the ensemble / configuration repository.

Missing from the original specification is a generic query function that retrieves arbitrary document (that is, XML) fragments. Such a function has not been implemented for two reasons:

- a. It is considered a low priority for users at the time of writing.
- b. It is not clear that different (XML) databases have a consistent return format for arbitrary XPath queries. For example, it has been observed that different versions of the eXist database [11] produce different forms of output for the same XPath query.

Furthermore, as the understanding within the Working Group has grown, the value of bespoke, domain-specific functionalities – that encapsulate the user interaction with the service and simplify the security of an implementation – has been realised.

2. The web service interface that has been defined is deliberately lightweight. This implies that client-side logic is required to implement complex queries. For example, it is the responsibility of the client to aggregate query results from multiple web service instances.

This is not considered an issue and the Working Group propose to introduce additional, domain-specific functionality, as the need arises.

3. The query retrieval interface supports the return of a subset of the matches found. This may be regarded as an analogue of an Internet Search Engine, in which search results are returned 10 at a time. There is a very small chance that a user requesting, for example, two subsequent sets of ten results (that is, results 1—10 followed by results 11—20) would not receive the first 20 results. This is because the service is stateless and one is not able to confirm that a particular database will return the same ordering of results if the same query is executed twice in succession. In fact, this same problem is encountered by Internet Search Engines, and for the same reason. The Working Group decided to note the potential issue, though not take any further action at the time of writing.
4. A number of simple client tools have been created as part of the development effort. These tools are lightweight and well-suited to testing and demonstration purposes. However, the existing clients are not at production level and require the user to have an understanding of the XPath query syntax. This is something that the Working Group believes could potentially hinder the uptake of the technology and suggests the importance of investing effort to develop a user-friendly and functional client.

4 Conclusions and further work

In May 2006, the ILDG published Version 1.0 of the specification for the ILDG Metadata Catalogue Web Service. At the time of writing four collaborations have completed implementations of the specification. These are JLDG, LQCD, UKQCD, and USQCD. In addition, one other collaboration – CSSM, Australia – are in the process of developing an implementation.

Four lightweight clients have also been developed to interface with the service and help to test individual implementations of the specification. These include two web clients from LQCD and USQCD.

Looking to the future, there are a number of important tasks that need to be addressed by the Working Group:

- ILDG need to develop a fully functioning client application that is tailored to a typical user. For example, the client should not require any prerequisite knowledge of XPath. The existing QCDgrid Browser has been nominated as an appropriate client. To this end, the QCDgrid project team have begun work on a design for a modified browser that fulfils the requirements of the ILDG and has functionalities that are specific to UKQCD removed.
- ILDG need to publicise the output of the working group to date, targeting both Lattice QCD and the wider scientific community. This will accelerate uptake of the technology by real users and help to maintain the current momentum within the Working Group.
- ILDG need to help and support other collaborations as they develop an implementation of the service. This support may include sharing of existing code – for example the UKQCD implementation can be easily and quickly tailored for any XML database that supports the XMLDB specification.
- Additional functionalities may need to be added to the specification. For example, the LQCD local user service provides a document validation operation, which allows a user to validate metadata against the corresponding QCDML schema, without actually submitting data to the catalogue.

Looking to the future the next task for the ILDG is to begin work on a specification for sharing metadata. It is important for the working group to review the strengths and weaknesses of the process to date, to help maintain and improve on the distributed working model that is employed.

In addition to the above, several other tasks must be completed by the QCDgrid team specifically:

- The UKQCD implementation of the web service must be deployed onto the production UKQCD Grid. This is expected to happen before the end of June 2006.
- Work Package 4 of the QCDgrid project plan needs to be revised to reflect the project's contribution to the Working Group's plan for the data sharing challenge.

References

- [1] D.J. Byrne, M.G. Beckett [*Work Package 3.3: QCDGrid2 Metadata Catalogue Web Service Specification*](#), QCDgrid Project Deliverable QCDGRID2-MDC-WS-SPEC_1.0 (November 2005).
- [2] D.J. Byrne [*Work Package 3.4: Metadata Catalogue Web Service Design*](#), QCDgrid Project Deliverable QCDGRID_MDC_WS_DESIGN (January 2006).
- [3] D.J. Byrne, M. G. Beckett [*ILDG MDC Web Service Test Suite \(Version 1.0\)*](#), QCDgrid Project/ILDG Deliverable (May 2006).
- [4] ILDG, International Lattice Data Grid home page, website <http://www.lqcd.org/ildg/tiki-index.php> (August 2005).
- [5] ILDG, [*ILDG MDC Web Service WSDL Definition \(Version 1.0\)*](#), ILDG Deliverable (May 2006).
- [6] ILDG, [*The 8th ILDG Workshop*](#), Meeting via VRVS (May 2006).
- [7] D. Pleiter, [*ILDG MDC-WS Operations \(Version 1.0\)*](#), ILDG Deliverable (May 2006).
- [8] QCDgrid: Probing the building blocks of matter with the power of the Grid, QCDgrid Project Homepage available on-line at <http://www.gridpp.ac.uk/qcdgrid/> (March 2006).
- [9] QCDgrid Project, [*GridPP Log Book – QCDgrid quarterly report, March 2006 \(Q1/06\)*](#), QCDgrid Project Deliverable (March 2006).
- [10] QCDgrid Project, *QCDgrid*, software application available on-line at <http://forge.nesc.ac.uk/projects/qcdgrid/> (February 2006).
- [11] SourceForge, Exist Version 1.0b, an Open Source, native XML database, available from <http://exist.sourceforge.net/> (April 2006).