FERMILAB-PUB-07-811-T

PROCEEDINGS OF SCIENCE

# Towards an interoperable International Lattice Datagrid

**Paul Coddington, Shunde Zhang**
*School of Computer Science, University of Adelaide, Adelaide, 5005, Australia*
*E-mail:* paulc@cs.adelaide.edu.au shunde.zhang@adelaide.edu.au

**Noriyoshi Ishii, Mitsuhisa Sato**
*Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan*
*E-mail:* ishiin@ccs.tsukuba.ac.jp, msato@cs.tsukuba.ac.jp

**David Melkumyan, Dirk Pleiter***
*Deutsches Elektronen-Synchrotron DESY, 15738 Zeuthen, Germany*
*E-mail:* dirk.pleiter@desy.de, david.melkumyan@desy.de

**George Beckett, Radoslaw Ostrowski**
*EPCC, School of Physics, University of Edinburgh, Edinburgh EH9 3JZ, Scotland*
*E-mail:* george.beckett@ed.ac.uk radek@epcc.ed.ac.uk

**James Simone**
*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
*E-mail:* simone@fnal.gov

**Balint Joó, Chip Watson**
*Jefferson Lab, Newport News, VA 23606, USA*
*E-mail:* bjoo@jlab.org, watson@jlab.org

**ILDG Middleware Working Group**

The International Lattice Datagrid (ILDG) is a federation of several regional grids. Since most of these grids have reached production level, an increasing number of lattice scientists start to benefit from this new research infrastructure. The ILDG Middleware Working Group has the task of specifying the ILDG middleware such that interoperability among the different grids is achieved. In this paper we will present the architecture of the ILDG middleware and describe what has actually been achieved in recent years. Particular focus is given to interoperability and security issues. We will conclude with a short overview on issues which we plan to address in the near future.

*The XXV International Symposium on Lattice Field Theory*
*July 30 - August 4 2007*
*Regensburg, Germany*

*Speaker.

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike Licence. 

arXiv:0711.4852v1 [hep-lat] 30 Nov 2007

## 1. Introduction

The goal of the International Lattice Datagrid (ILDG) is to establish an international grid infrastructure which provides the means for longterm storage and global sharing of data produced during compute-intensive lattice Quantum Chromodynamics (QCD) simulations.

For data sharing and data procurement the availability of metadata is a key requirement. With common standards being developed, implemented and used in practice it becomes possible to make metadata publicly available such that all pieces of information have a well-defined semantic meaning. Within ILDG, the agreed standard for describing scientific metadata is encapsulated in a family of XML schema, called QCDml. For an overview and status report on the definition of QCDml see [1].

The lattice QCD community, like many other research communities, is not in the position to globally develop, deploy and operate a common middleware stack. Instead, application specific software has to be built on top of existing middleware frameworks (e.g., gLite or the Globus Toolkit) and grid infrastructures (like LCG/EGEE or the Open Science Grid). On the global level this requires the realisation of a concept of grid-of-grids. Making the various grid services interoperable continues to a major challenge. Two strategies are employed to face the interoperability challenge. Firstly, common grid standards are adopted, whenever possible. An example is the Storage Resource Manager protocol (SRM) [2], a protocol evolving to an open standard for grid middleware to communicate with site specific storage fabrics. Secondly, interface services are defined and implemented. These services are implemented as web-services for which a set of operations has been standardised by a Webservice Description Language (WSDL) document, a behavioural specification and the specification of test suites. For instance, the ILDG file catalogue interface service allows to interface to different file catalogues which are being used, e.g. the LCG File Catalogue (LFC) or the Globus Replica Location Service (RLS).

The following regional grids have been put into operation: CSSM (Australia) [3], Latfor Datagrid (LDG) (continental Europe) [4], JLDG (Japan) [5], UKQCD grid (UK) [6], and USQCD (US) [7].

## 2. ILDG middleware functionality

The currently deployed services allow to query for and to download metadata as well as to download the data.[1] Typically, a user will start by performing a query for available ensembles. The ensembles are described by ensemble XML documents stored in a Metadata Catalogue (MDC). Each ensemble can be addressed by a globally unique URI, the so-called `markovChainURI`. The unique identifier for each configuration and configuration metadata document is a logical filename, `dataLFN`. (See [1] for further details on the `markovChainURI` and `dataLFN`.) To actually download configurations that belong to a particular ensemble the following operations have to be performed:

1. Query the metadata catalogue (MDC) for all configurations which have a particular ensemble URI in common. The MDC will return a list of logical file names.

[1] Note that within the regional grids more operations, in particular data upload operations, are available.

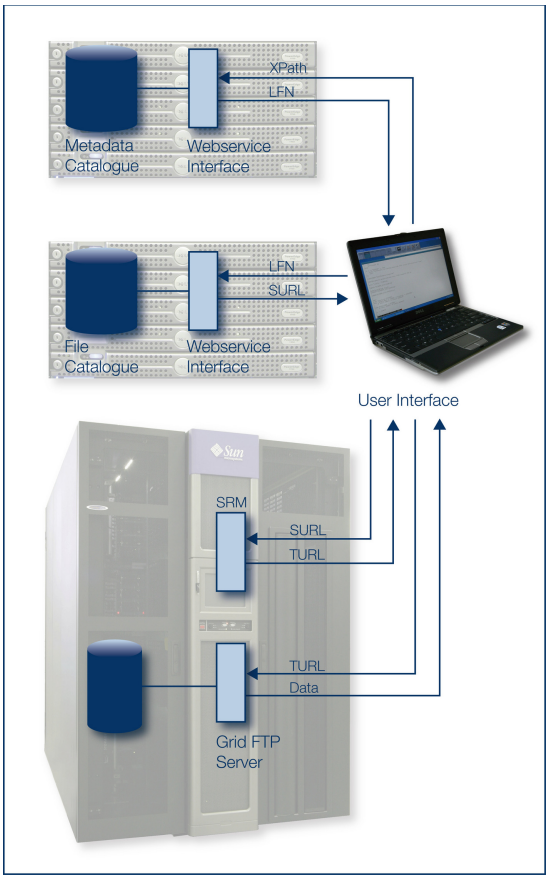


**Figure 1:** Simplified overview on the middleware architecture. On the left-hand side the different services metadata catalogue (MDC), file catalogue (FC) and Storage Element (SE) are shown (top to bottom) which are deployed in all of the ILDG regional grids. The arrows indicate the operations initiated by the client.

2. Query the file catalogue (FC) to identify all replicas of a file identified by its logical filename `dataLFN`. The FC will return a list of Site Uniform Resource Locators (SURL).[2]
3. If the scheme of the SURL is equal to 'srm' a Storage Resource Manager will have to be queried to obtain the actual transfer URL (TURL). Otherwise, SURL and TURL are identical.
4. Download the data from a Storage Element (SE) using the TURL.

This sequence of operations and the involved services are schematically shown in Fig. 1. Note that a user does not have to care about each of these steps as client tools will typically process a sequence of these operations by default.

Operations needed to upload and modify data are currently not standardised within ILDG and are hence regional grid specific.

[2]An SURL is a kind of Uniform Resource Identifier (URI).

## 3. Middleware services

The *Metadata Catalogue* (MDC) is a database which stores the ensemble and configuration XML documents. Each of the regional grids has deployed such a service together with a webservice interface that provides at least the following operations standardised by ILDG:

**`doEnsembleURIQuery`** Queries the MDC for ensemble metadata documents which match a given XPath expression. For each matching document the corresponding ensemble URI is returned.

**`doConfigurationLFNQuery`** Similar to `doEnsembleURIQuery` but for configuration metadata documents. A list of logical file names (LFN) is returned.

**`getEnsembleMetadata`** Return ensemble metadata document for a given ensemble URI.

**`getConfigurationMetadata`** Return configuration metadata document for a given LFN.

Several web-portals have been deployed by the regional grids which allow to query all the MDCs. Furthermore, EPCC has developed the ILDG Browser Client. This java-based browser allows to construct an XPath query in an intuitive way and to submit this query to any of the MDCs [6].

The *File Catalogue* (FC) stores for each logical file name at least one Site Uniform Resource Locator (SURL). Such an SURL identifies a copy of a data file (e.g. gauge configuration) on any of the storage elements. For each data file more than one replica may exist. Different file catalogues are being used by the various regional grids. Therefore, a webservice interface has been defined and an interface service has been implemented as a joined effort by DESY and EPCC. The interface service receives queries via the standardised webservice interface, forwards them to the regional grid specific FC and finally returns its response via the webservice interface back to the client. So far only one interface operation is available. The `getURL` operation takes a list of LFNs on input and returns for each of them all registered SURLs. All regional grids have deployed and tested this service. A complication of the FC interface service results from authentication being mandatory for some file catalogues. If this is the case then the interface service has to able to act on behalf of the user. Or to put it in different words: The user client has to delegate its credentials to the interface service. (See below for more details on the credential delegation service.)

The real data is stored on so-called Storage Elements (SE) where the data can be found using its SURL. A large variety of SEs configurations exist. A simple SE may consist of a single server with an attached disk array. A large SE may comprise a whole set of file servers which are used to stage the data stored in a tape back-end system (HSM). Typically, the latter kind of SE will provide a Storage Resource Manager (SRM) interface [2].[3] The SRM protocol allows the client and server to negotiate the transport protocol and to notify the client about the server which will eventually provide the data. This feature gives the SE the freedom to dynamically assign a file server on which the data is being staged depending, e.g., on its current load. For final data transport the ILDG middleware architecture allows for GridFTP and http. This flexibility at the server side requires somewhat fatter clients as ILDG download clients must be able to communicate with an SRM interface and support all transport protocols.

[3] SRM v1 is currently used, but it is expected that the SEs will be upgraded to SRM v2.2 in the near future.

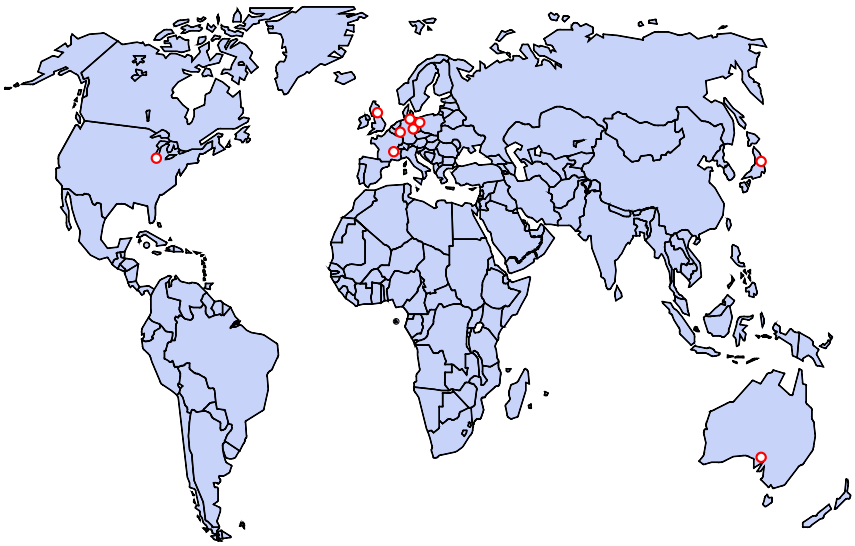

**Figure 2:** ILDG storage elements currently in operation.

We started to test interoperability of the services and verified that from any of the regional grids download operations from SEs operated by any of the regional grids is indeed possible. A prototype client, called `ildg-get`, has been implemented. This allows to download metadata from any of the metadata catalogues, to query the file catalogues for replica of a particular file and, finally, to download the file from an SE.

## 4. Security and VO management

The security infrastructure employed for ILDG is common to all major grid middlewares. Before a secure operation is started user and service perform a mutual *authentication*. Both sides have to have a pair of private key and public X.509 certificate. The private key is confidential and not transmitted via the network. The public certificate is signed by a trusted Certificate Authority (CA). The server (client) can check whether the private key/public certificate pair matches and whether the user (host) certificate has been signed by a trusted CA. The later is done by comparing the certificate with the root certificates published by the CAs and which have to be installed both on the server and the client.

Trusting the CA means that we rely on the CAs to fulfil certain standards when providing certificates to users, e.g. they have to ensure that the user did identify himself. To reduce the risk of user (host) credentials being compromised the life-time of certificates are limited. Furthermore, the CAs regularly publish lists of revoked certificates, so-called Certificate Revocation Lists (CRL). Both, client and server always check that the CRL is not out-of-date. In ILDG it has been agreed to trust certificates issued by any of the Certificate Authorities (CA) of the International Grid Trust Federation (IGTF) [9].

To actually obtain access to any of the ILDG resources a user has to join the *Virtual Organisation* (VO) ILDG. Registration with and management of the VO is done via the *Virtual Organization Membership Registration Service* (VOMRS) [8]. A user which wants to join to the VO has to perform the following steps:

- Load certificate into a web-browser and establish an authenticated connection to the VOMRS service.
- In Phase I, the user fills in a registration form and nominates a regional grid (institution) plus representative, to approve their application.
- After email address has been verified the candidate is invited to proceed to phase II of the registration procedure during which he/she can apply for joining a group and has to accept the VO Usage Rules as defined in the ILDG VO Policy document [10].

There are 2 or 3 representatives for each of the regional grids. They will decide only on those applications coming from the corresponding regional grid. They are responsible for approving requests according to the VO Policy document which, e.g., means that they have to verify that the applicant is a scientist performing research in lattice QCD.

The VOMRS service propagates VO membership information to a VOMS service from where this information can be retrieved by each of the regional grids. All regional grids automatically synchronise their local user mappings with this central VO user management system. Once a representative approved the application for VO membership this new user will automatically become known to all services in all the regional grids.

For each of the regional grids a corresponding group has been defined. The information about group membership can be used by the regional grid services for authorisation, e.g. to restrict access to data to members of that particular regional grid. Only the owner and managers of a group are allowed to add VO members to that particular group.

Particular care is needed when credentials are delegated by the client to an interface service. For the FC interface service we use a credential delegation service which has become part of the gLite middleware [11]. If a client requests a credential to be delegated the server generates a private key and a certificate request. The latter is sent to the client for being signed. After this public certificate has been returned, the server can use it to authenticate on behalf of the user with other servers. Note that the private key is never transmitted via the (potentially insecure) network. For security reasons the maximum time of validity of the delegated credentials is limited to 12 hours. Once the credential has expired the procedure of credential delegation has to be repeated.

## 5. Conclusions and outlook

Significant progress has been made since ILDG was initially proposed in 2001. Various regional grids have been deployed for production use. This infrastructure is being used by lattice researchers all around the world to make an increasing amount of data available. An overview on the current usage has been given by C. DeTar at this conference [12].

Admittedly, a user who wants to gain access to the grid for the first time has to make substantial efforts. The user has to

- Install grid client software.
- Identify a Certificate Authority (or one of its Registration Authorities) and request a certificate.
- Join the Virtual Organisation.

We nevertheless observe that these efforts start to pay-off for an increasing number of people being involved in lattice QCD research. In particular large physics collaborations where tasks of generating gauge field configurations and calculating observables on these configurations have been divided between members distributed over different sites and countries start to benefit from this new research infrastructure.

Still a lot of work needs to be done to improve the usability of this infrastructure. Issues which the ILDG Middleware Working Group plans to address in the near future include:

- More efforts are needed to improve functionality and operation of all regional grids.
- Reliable operation of all regional grid services should be improved by putting a monitoring system in place.
- The available functionality of the services standardised by ILDG should be extended, in particular, to allow for replication of data beyond the regional grid boundaries. This would be relevant for physics collaborations which comprise of members in different regional grids. Furthermore, it would allow to optimise access time to data which would otherwise only be available from storage elements where the bandwidth of the network connecting client and server is poor.

## Acknowledgements

We are indebted to all institutions and people that helped to build-up the regional grid infrastructures and provide their resources. DP acknowledges the support from C. Iezzi for preparing this presentation.

## References

[1] T. Yoshié et al. [ILDG Metadata Working Group], *Marking up lattice QCD configurations and ensembles*, PoS(LATTICE 2007)161.

[2] *Storage Resource Manager protocol rev. 2.2* (http://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.pdf).

[3] CSSM grid (http://cssm.sasr.edu.au/ildg/).

[4] LatFor Datagrid (http://www-zeuthen.desy.de/latfor/ldg/).

[5] Japan Lattice Datagrid (http://www.jldg.org/lqa/index.html).

[6] QCDgrid/DiGS (http://www.gridpp.ac.uk/qcdgrid/).

[7] USQCD grid (http://usqcd.jlab.org/).

[8] Registration site for the ILDG virtual organisation: https://grid-voms.desy.de:8443/vo/ildg/vomrs/.

[9] International Grid Trust Federation (http://www.gridpma.org).

[10] ILDG Middleware Working Group, *ILDG VO Policy* (Rev. 1.0).

[11] A. McNab and S. Kaushal, *The GridSite Proxy Delegation Service*, UK e-Science All Hands Conference, Nottingham, September 2006.

[12] C. DeTar, *Sharing Lattices Throughout the World: An ILDG Status Report*, PoS(LATTICE 2007)009.