# EGI-InSPIRE

## STANDARDS ROADMAP

### EU DELIVERABLE: D2.5

| | |
|---|---|
| Document identifier: | EGI-D2.5-206-V4.doc |
| Date: | **24/11/2010** |
| Activity: | **NA2** |
| Lead Partner: | **EGI.eu** |
| Document Status: | **FINAL** |
| Dissemination Level: | **PUBLIC** |
| Document Link: | https://documents.egi.eu/document/206 |

Abstract

This document contains an overview of the relevant standards activity taking place in the EGI environment, both internally within the operational tools and through external software providers as described in the UMD roadmap.

## II. DELIVERY SLIP

| | Name | Partner/Activity | Date |
|---|---|---|---|
| **From** | Sergio Andreozzi | EGI.eu | 12-10-2010 |
| **Reviewed by** | **Moderator:** Gergely Sipos<br>**Reviewers:** Morris Riedel (FZJ, Juelich, DE), Steve Crouch (University of Southampton, UK) | EGI.eu | 19-10-2010 |
| **Approved by** | **AMB & PMB** | | 24-11-2010 |

## III. DOCUMENT LOG

| Issue | Date | Comment | Author/Partner |
|---|---|---|---|
| 1 | 12-10-2010 | Full draft | Sergio Andreozzi/EGI.eu |
| 2 | 05-11-2010 | Updates based on reviewers'comments | Sergio Andreozzi/EGI.eu |
| 3 | 11-10-2010 | Final version | Sergio Andreozzi/EGI.eu |

## IV. APPLICATION AREA

This document is a formal deliverable for the European Commission, applicable to all members of the EGI-InSPIRE project, beneficiaries and Joint Research Unit members, as well as its collaborating projects.

## V. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE "Document Management Procedure" will be followed: https://wiki.egi.eu/wiki/Procedures

## VI. TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/.

## VII. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed − both for coordinating the infrastructure and for delivering integrated services that cross national borders. The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting 'grids' of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example within the ESFRI projects. Additional support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today's production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI.eu, coordinator of EGI-InSPIRE, brings together partner institutions established within the community to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the community.

The production infrastructure supports Virtual Research Communities (VRCs) − structured international user communities − that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

# VIII. EXECUTIVE SUMMARY

The realisation of the EGI vision requires the ability to cross both organisational and technical boundaries. Such an aspect is usually referred as interoperability, i.e., the ability of systems, people and organisations to provide services to and accept services from other systems, people and organisations and to use the services so exchanged to enable them to operate effectively together.

Reaching interoperability amongst organisations and technologies is a long-term activity, which requires reaching consensus through compromises and reworking/rebuilding systems or procedures according to them. Interoperability can be addressed at different levels leading to the identifications of different interoperability types. In our context, we identify three main types of interoperability: 1) strategic, 2) operational and 3) technical.

For each type of interoperability, proper actions should be taken in order to enable it. At the technical level, there are two main approaches: adapter-based and standards-based interoperability. The former envision that adapters between interacting parties are developed to translate the specific requests from the client side to the equivalent format and protocol on the service side. The latter envisions the definition of a common interface and message format as an open standard. In this case, the parties are expected to refactor or appropriately extend their systems in order to comply with the standard specification.  It is widely recognised that standardisation is one of the key facilitators for interoperability of networks, services and equipment [R2].

This document focuses on standards-based technical interoperability, i.e., the interoperability amongst systems participating in EGI (e.g., computing clusters, storage systems) through the adoption of open standards. By systems, we mainly consider the software abstraction layer (middleware) needed to expose the functional and operational interfaces outside the organisational boundaries together with the security mechanisms needed by the EGI infrastructure. Open standards are mapped into the UMD capabilities [R3]. This document represents an evolving roadmap that will officially updated and published every twelve months, while being always open for contributions.

## TABLE OF CONTENTS

# 1 INTRODUCTION

EGI's main goal is to provide a secure integrated federated computing infrastructure constructed from national and domain specific resource providers. Such an infrastructure should be open to scientists and researchers from Europe and worldwide to support their day-to-day work. Different computing models should be supported as well as access to different types of distributed resources (high-throughput, high-performance, desktop, virtualised, etc.) linked to physically remote data stores.

The realisation of the EGI vision requires the ability to cross both organisational and technical boundaries. Such an aspect is usually referred to as interoperability. In the last decade, many definitions for interoperability were proposed. Given the complexity of our context, we favour the following broad definition inspired by activity in the military area [R1]: "Interoperability is the ability of systems, people and organisations to provide services to and accept services from other systems, people and organisations and to use the services so exchanged to enable them to operate effectively together".

Reaching interoperability amongst organisations and technologies is a long-term activity, which requires reaching consensus through compromises and refactoring/rebuilding systems or procedures according to them. Interoperability can be addressed at different levels leading to the identifications of different interoperability types. In our context, we identify three main types of interoperability: 1) strategic, 2) operational and 3) technical.

For each type of interoperability, proper actions should be taken in order to enable it. At the technical level, there are two main approaches: adapter-based and standards-based interoperability. The former envision that adapters between interacting parties are build to translate the specific requests from one side to the equivalent format and protocol on the other side. The latter envisions the definition of a common interface and message format as an open standard. In the case of adapters, it is known that such an approach raises the issue of maintenance overheads due to the necessary transformation logic, including, in some cases, a loss of functionality and/or semantic correctness. In the case of standards, the parties are expected to refactor or appropriately extend their systems in order to comply with the common specification. In addition, it is widely recognised that standardisation is one of the key facilitators for interoperability of networks, services and equipment [R2].

This document focuses on standards-based technical interoperability, i.e., the interoperability amongst the systems participating in EGI (e.g., computing clusters, storage systems) through the adoption of open standards. By systems, we mainly consider the software abstraction layer (middleware) needed to expose the functional and operational interfaces outside the organisational boundaries together with the security mechanisms needed by the EGI infrastructure. Open standards are mapped into the UMD capabilities [R3]. This document represents an evolving roadmap that will officially updated and published every twelve months, while being always open for contributions.

# 2 INTEROPERABILITY AND STANDARDS

As anticipated in the previous section, interoperability is an intrinsic requirement for organisations and systems that need to operate effectively together. We can argue that interoperability is not a merely technical aspect, but it also affects the strategic and operational level of organisation's interactions. In the strategic area, it is important to engage organisations and funding agencies to evolve policies that raise the awareness of the importance of e-infrastructures interoperability thus influencing strategic agendas or funding programs. In the operational area, interoperability can be achieved by defining common policies and procedures regulating the way different organisations or group of persons interact. In the technical area, the adoption of open standards can facilitate interoperability and avoid vendor lock-in by infrastructure providers while providing users with more choices of service providers and less overhead in integrating/maintaining their applications into the e-infrastructure (see Figure 1).

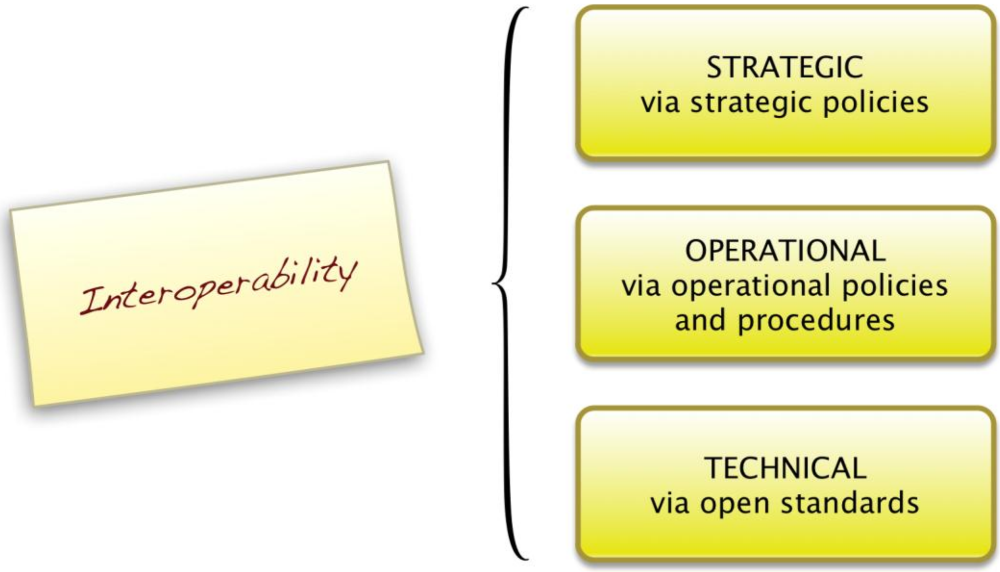


**Figure 1 - The three types of interoperability**

From the user perspective, interoperability is achieved when the expectations of the user to perform computations or store/manipulate data into an integrated e-infrastructure is met withouth being forced to use one particular technology or approach. Therefore, from the user perspective, the fundamental value of interoperability is enhancing users' choice and their ability to scale-up the use of resources when required.

Open standards are widely recognised to facilitate interoperability; nevertheless they are not strictly sufficient to enable it. Previous experience has shown that although standards can be present, their incomplete implementation or the degree of ambiguities and flexibility in the specification can lead to non-interoperable systems. Therefore, the presence of conformance test suites covering a large number of realistic use cases is essential to verify the interoperability of the deployed standard-compliant implementations.

When referring to the term standard, we intend a “document, established by consensus and approved by a recognized body, which provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context”.

Standards are open when they have the following four properties [R2]: 1) *control*: the evolution of the specification should be set in a transparent process open to all interested contributors; 2) *completeness*: the technical requirements of the solution should be specified completely enough to guarantee full interoperability; 3) *compliance*: there is a substantial standard-compliant offering promoted by proponents of the standard; 4) *cost*: fair reasonable and non-discriminatory access is provided to all implementers.

# 3 STANDARDS RELEVANT TO EGI

This section is dedicated to identify the standards that are relevant to EGI together with the Standards Developing Organisations (SDOs) from which they originate. Standards are matched against the abstract capabilities that a middleware for DCIs should provide according to the UMD Roadmap [R3].

## *3.1 Standard Bodies*

In this section, we list the main Standards Developing Organisations (SDOs) that produce standards output useful to enable an integrated and federated e-Infrastructure and the engagement of the EGI community with the bodies.

### 3.1.1 W3C

The World Wide Web Consortium (W3C) [R4] is an international community where Member organizations, a full-time staff, and the public work together to develop Web standards. Led by Web inventor Tim Berners-Lee and CEO Jeffrey Jaffe, W3C's mission is to lead the Web to its full potential by developing protocols and guidelines that ensure Web long-term growth.

### 3.1.2 WS-I

The Web Services Interoperability Organization (WS-I) [R5] is an open industry organization chartered to establish Best Practices for Web services interoperability, for selected groups of Web services standards, across platforms, operating systems and programming languages. WS-I comprises a diverse community of Web services leaders from a wide range of companies and standards development organizations (SDOs). WS-I committees and working groups create Profiles and supporting Testing Tools based on Best Practices for selected sets of Web services standards. The Profiles and Testing Tools are available for use by the Web Services community to aid in developing and deploying interoperable Web services. Companies interested in helping to establish Best Practices for Web Services are encouraged to join WS-I.

### 3.1.3 IETF

The Internet Engineering Task Force (IETF) [R6] is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The actual technical work of the IETF is done in its working groups, which are organized by topic into several areas (e.g., routing, transport, security). Much of the work is handled via mailing lists. The IETF holds meetings three times per year.

### 3.1.4 OASIS

The Organization for the Advancement of Structured Information Standards (OASIS) [R7] is a not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society. The consortium produces standards for Web services, security, e-business, for both the public sector and for application-specific markets. OASIS is distinguished by its transparent governance and operating procedures. Members themselves set the OASIS technical agenda, using a lightweight process expressly designed to promote industry consensus and unite disparate efforts.

### 3.1.5 OGF

The Open Grid Forum (OGF) [R8] is an open community committed to driving the rapid evolution and adoption of applied distributed computing. Applied Distributed Computing is critical to developing new, innovative and scalable applications and infrastructures that are essential to productivity in the enterprise and within the science community. OGF accomplishes its work through open forums that build the community, explore trends, share best practices and consolidate these best practices into standards.

### 3.1.6 DMTF

The Distributed Management Task Force (DMTF) [R9] enables more effective management of millions of IT systems worldwide by bringing the IT industry together to collaborate on the development, validation and promotion of systems management standards. The group spans the industry with 160 member companies and organizations, and more than 4,000 active participants crossing 43 countries. The DMTF board of directors is led by 15 innovative, industry-leading technology companies. With this deep and broad reach, DMTF creates standards that enable interoperable IT management. DMTF management standards are critical to enabling management interoperability among multi-vendor systems, tools and solutions within the enterprise.

### 3.1.7 IEEE

IEEE [R11] is the world’s largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE and its members inspire a global community through IEEE's highly cited publications, conferences, technology standards, and professional and educational activities. IEEE is led by a diverse body of elected and appointed volunteer members. The governance structure includes boards for operational areas as well as bodies representing members in the 45 societies and technical councils and ten worldwide geographic regions.

## *3.2 Standards vs. Capabilities*

In this section, we present the standards relevant to EGI [R10] categorized vie necessary capabilities of technologies required by the infrastructure that have been identified in the UMD Roadmap [R3]. These capabilities are classified into three areas (see Figure 2):

- Functional: a functional capability would be generally consumed by an end-user community
- Security: a security capability is expected to cut across all functional and managerial components
- Operational: an operational capability is provided to help with the operation of the functional capability within the production infrastructure.

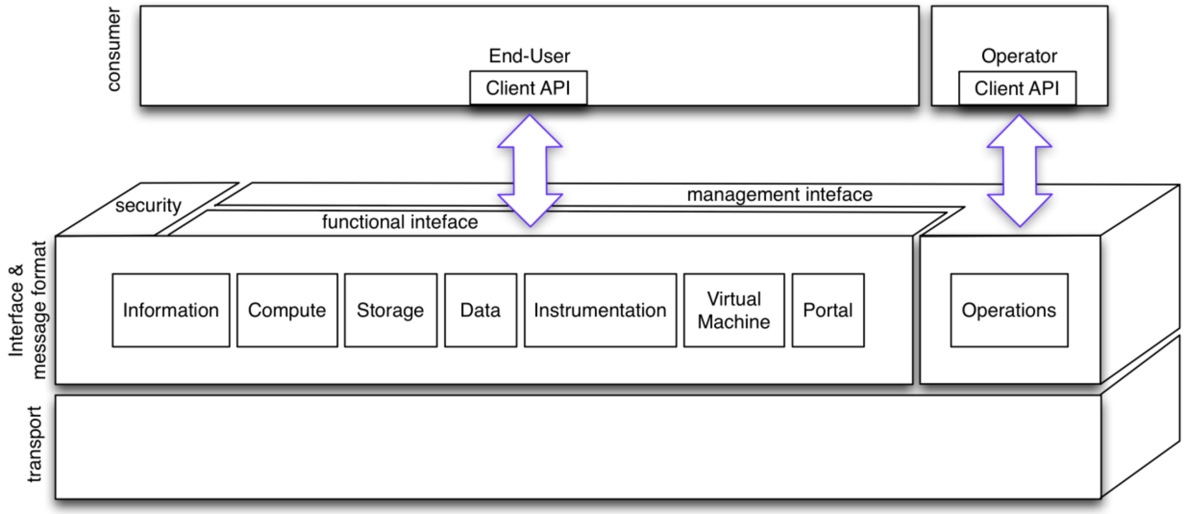


**Figure 2 - Capability stack in the UMD roadmap**

For each standard, the development phase is also specified by using the following classification also summarized in Figure 3:

1. PRE (Preparation): the needs for a standard has been identified, partnering among stakeholders is ongoing as well as the identification of an appropriate SDO
2. DEV (Development): a working group within an SDO has accepted to work on a standard and the development of the specification, potentially starting from use cases, is started
3. IMP (Implementation): a specification has been proposed and ratified, implementations of the standard is started by at least one EGI technology provider
4. DEP (Deployment): deployment in production systems of components complying with the standard is started by at least one organisation providing services to EGI
5. USE (Use): the standards-based functionalities of components are being used as primary functionalities replacing the legacy ones in day-to-day activity by at least one virtual community


**Figure 3 - Phases in Standard Development**

Concerning the transport layer (see Figure 2), the EGI infrastructure relies on a common set of standards that are widely adopted (e.g., HTTP v1.1, XML, XML Schema, WS-Interoperability Basic Profile). This document focuses mainly on higher-level standards which are specific to the DCI capabilities being essentially relevant for message formats, interfaces and API.

**Table 1 - Capability vs. Standard**

| Capability | Standard | SDO | Phase |
|---|---|---|---|
| Information.Messaging | JMS 1.1 | JCP | USE |
| Information.Messaging | AMQP 1.0 | AMQP WG | IMP |
| Information.Discovery | LDAPv3 | IETF | USE |
| Information.Model | GLUE 2.0 | OGF | DEP |
| Compute.JobExecution | BES, JSDL, HPC-BP, HPC File Staging Profile | OGF | DEP |
| Compute.JobExecution | PGI | OGF | DEV |
| Compute.WorkflowExecution | WS-BPEL 2.0 | OASIS | DEP |
| Compute.WorkflowExecution | DCI-Federation | OGF | DEV |
| Compute.ParallelJobExecution | MPI | MPI-Forum | USE |
| Compute.ParallelJobExecution | OpenMP | OpenMP ARB | USE |
| Compute.ParallelJobExecution | HPC SPMD | OGF | DEV |
| Compute.JobScheduling | | | |
| Storage.Management | SRM v2.2 | OGF | USE |
| Storage.FileTransfer | GridFTP | OGF | USE |
| Storage.FileTransferScheduling | DMI | OGF | DEV |
| Storage.FileAccess | NSF 4.1 | IETF | IMP |
| Storage.FileEncryption | | | |
| Data.Access | WS-DAIR, WS-DAIX | OGF | IMP |
| Data.MetadataCatalogue | | | |
| Instrumentation.Management | | | |
| Security.Authentication | X.509 + RFC3820 | IETF | USE |
| Security.Authentication | SAML 2.0 | OASIS | DEP |
| Security.Authentication | OpenID | OIDF | IMP |
| Security.Authorisation | SAML 2.0, XACML 2.0 | OASIS | DEP |
| Security.CredentialManagement | | | |
| Security.AttributeAuthority | SAML 2.0 | OASIS | DEP |
| VirtualMachine.Management | OCCI | OGF | DEV |
| VirtualMachine.ImageFormat | OVF | DMTF | IMP |
| VirtualMachine.ImageDistribution | | | |
| Operations.Monitoring | | | |
| Operations.Accounting | UR | OGF | IMP |
| Operations.Accounting | RUS | OGF | DEV |
| ClientAPI | SAGA | OGF | IMP |

### 3.2.1 Information.Messaging

Within distributed systems, a message 'bus' provides a reliable mechanism for data items to be sent between producers and (multiple) consumers. Such a capability, once established, can be reused by many different software services.

#### 3.2.1.1 Available Standards

In this area, the Java Messaging Service (JMS) API is a standard from the Java Community Process (JCP) that allows application components based on the Java 2 Platform, Enterprise Edition (J2EE) to create, send, receive, and read messages. It allows the communication between different components of a distributed application to be loosely coupled, reliable, and asynchronous. AMQP (Advanced Message Queuing Protocol) is emerging as de-facto standard in this area.

### 3.2.2 Information.Discovery

The Information discovery capability relates to the ability of locating unknown resources or services that are part of the infrastructure, possibly to satisfy a set of requirements. Such a capability implies the need for a common interface for running the discovery process.

#### 3.2.2.1 Available Standards

LDAP v3 is an IETF specification of a protocol, a query language and a data model designed to provide access to directories supporting the directories supporting the X.500 models.

### 3.2.3 Information.Model

The Information model capability relates to the ability of modelling resources within EGI based on community-accepted definitions.

#### 3.2.3.1 Available Standards

GLUE 2.0 is an OGF specification for an information model for Grid entities described using the natural language and UML Class Diagrams. As a conceptual model, it is designed to be independent from the concrete data models adopted for its implementation. Rendering to concrete data models such XML Schema, LDAP Schema and SQL Data Definition Language (DDL) are provided in separate specifications.

### 3.2.4 Compute.JobExecution

The compute capability relates to the ability to describe, submit, manage and monitor a work item on a specific site submitted for either queued batch or interactive execution. This capability should provide also detailed information about every step of the work item lifecycle, from the submission to unrecoverable failures or success.

#### 3.2.4.1 Available Standards

There are a number of different proprietary interfaces currently in production that provide the ability to describe, submit and manage an interactive or batch work item on a specific site. Activity within OGF in recent years has led to specifications in this area: Job Submission Description Language (JSDL), the Basic Execution Service (BES) and the High Performance Computing Basic Profile (HPC-BP) specifications. These specifications, the experiences derived from them and from other proprietary interfaces are forming the basis of ongoing activity within the Production Grid Infrastructure Working

Group. It is expected that the output from this activity will eventually lead to the interfaces that will be supported by EGI.

### 3.2.5 Compute.WorkflowExecution

The workflow execution capability relates to the ability to define, initiate, manage and monitor a workflow. Such a capability can be provided at the user or user community level and does not necessarily needs to be part of the core infrastructure. Nevertheless, various workflow systems may have requirements that need to be supported within the core infrastructure.

#### 3.2.5.1 Available Standards

In this area, there are a number of proposals, both proprietary and standards. The most popular standard is the Business Process Execution Language (BPEL), short for Web Services Business Process Execution Language (WS-BPEL) from OASIS. BPEL is an executable language for specifying actions within Business processes with Web services. Processes in BPEL export and import information by using Web service interfaces exclusively.

### 3.2.6 Compute.ParallelJobExecution

The parallel job execution capability relates to the ability to run parallel jobs in a Grid infrastructure. The infrastructure does not provide support at the programming level, but provides support for controlling the distribution of processes to physical machines within a cluster. The ability to have fine-grained control over the placement of processes for an MPI or OpenMP application is a key differential between this capability and a conventional batch job capability. This capability may require to extend simple job submission interface standards to handle a parallel execution of jobs.

#### 3.2.6.1 Available Standards

In this area, it is required that computing nodes supporting libraries for parallel job execution should properly discoverable via the information service. The required libraries are:

- MPI: Message Passing Interface 1.x
- MPI: Message Passing Interface 2.x
- OpenMP

Concerning GLUE 2.0, the property related to the support of parallel execution is already included in the class `ApplicationEnvironment`, property `ParallelSupport`. Regarding the submission of parallel jobs, in the set of OGF standards, the extention to JSDL called HPC SPMD was drafted. Nevertheless, requirements in this area are an essential part of the PGI standard endeavours potentially considered in the Compute.JobExecution capability.

### 3.2.7 Compute.JobScheduling

The compute Job scheduling capability refers to the ‘end-to-end’ service that can be delivered to a user in response to their request for a job to be run. This includes managing the selection of the most appropriate resource that meets the user’s requirements, the transfer of any files required as input or produced as output between their source or destination storage location and the selected computational resource, and the management of any data transfer or execution failures within the infrastructure. Such a capability also considers the lifecycle management of the job, from the preparation to the collection of results, including potential rescheduling to other resources upon failures, moreover accurate information about every step of the job lifecycle should be provided.

#### 3.2.7.1 Available Standards

In this context, there are no available standards being used in the Grid context. The OGF Distributed Computing Infrastructure Federation Working Group (DCIFED-WG) has planned to address this area, while some of the existing implementations of meta-schedulers support simpler interface (like OGF BES and JSDL).

### 3.2.8 Storage.Management

This capability relates to the ability of managing a storage resource, from simple systems like disk-servers to complex hierarchical systems.

#### 3.2.8.1 Available Standards

The Storage Resource Management (SRM) interface specification from OGF is a commonly adopted specification. However, there are ambiguities in the SRM interface and behaviour that need to be addressed and it has different levels of adoption within the compliant implementations.

### 3.2.9 Storage.FileTransfer

This capability related to the capacity of moving a file from one network location to another.

#### 3.2.9.1 Available Standards

In this context, there are many standards available. The most common used in Grid is GridFTP, but other options are possible (e.g., HTTP, BitTorrent).

### 3.2.10 Storage.FileTransferScheduling

This capability relates to the capacity of managing a transfer of files from the start to the completion.

#### 3.2.10.1 Available Standards

The main standard available in this context is the Data Movement Interface (DMI) from OGF.

### 3.2.11 Storage.FileAccess

This capability related to the ability of accessing parts of a networked file in read/write mode.

#### 3.2.11.1 Available Standards

POSIX (Portable Operating System Interface for UNIX) is the name of a family of related standards specified by the IEEE to define the application programming interface (API), along with shell and utilities interfaces for software compatible with variants of the Unix operating system, although the standard can apply to any operating system. One of the most important abstractions of the POSIX API is the file. A number of proprietary protocol definitions exist in order to provide remote read/write access via a POSIX-like interface. In particular, RFIO (Remote File I/O), DCAP (dCache Access Protocol) and XROOTD are common solutions being deployed in EGI. Another standard relevant in this area is NFS 4.1 from IETF.

### 3.2.12 Storage.FileEncryption

Sensitive data needs to be stored securely. Before being stored in a remote file store the file may need to be encrypted and then on retrieval de-encrypted before use. The capability should also provide solutions relating to the storage of the keys needed to perform these tasks.

#### 3.2.12.1 Available Standards

There are no available standards for this capability.

### 3.2.13 Data.Access

This capability relates to the ability of accessing over the Internet structured or semi-structured data in read/write mode.

#### 3.2.13.1 Available Standards

The WS-DAI set of standards from the OGF DAIS Working Group include specifications for accessing relational data, XML data and RDF data.

### 3.2.14 Data.MetadataCatalogue

The metadata catalogue is used to store and query information relating to the data (files, databases, etc.) stored within the production infrastructure. Functionalities include the ability to store and query information relating to the data item including, location, mapping of persistent storage identifiers to the locations of the stored data.

#### 3.2.14.1 Available Standards

There are no available standards for this capability.

### 3.2.15 Instrumentation.Management

Instruments are data sources frequently encountered within e-Infrastructures. As part of a distributed computing architecture providing remote access to manage and monitor these instruments is becoming increasingly important within some communities.

#### 3.2.15.1 Available Standards

There are no available standards for this capability.

### 3.2.16 Security.Authentication

This capability relates to the ability of providing authentication mechanisms for Grid users, machine and services. An authentication token that is strongly bound to an individual must be applied consistently across the software used within the production infrastructure. The authentication system must be capable of supporting a delegation model.

#### 3.2.16.1 Available Standards

IETF PKI and the IETF Proxy Certificate are two key pillars of the current Grid infrastructure to handle both authentication and delegation. Other standards to be considered are SAML from OASIS as a different approach to cross-organization authentication. OpenID provides federated identity using a text-based protocol, similar to SAML. Much more suitable for end-user Web-pages and RESTful Web services.

### 3.2.17 Security.Authorization

This capability relates to the ability of handling authorization aspects, making authorization decisions about the subject and the requested mode of access based upon combining information from a number of distinct sources.

#### 3.2.17.1 Available Standards

In this area, XACML from OASIS is a perfect coupling with SAML. It allows the definition of powerful security policies that can be configured in a way to match end-users attribute statements encoded in SAML assertions.

### 3.2.18 Security.CredentialManagement

This capability relates to the ability of providing an online credential repository that allows users to securely obtain credentials when and where needed. The interface should allow for obtaining, delegating and renewing authentication credentials by a client using a remote service.

#### 3.2.18.1 Available Standards

There are no available standards for this capability.

### 3.2.19 Security.AttributeAuthority

Resources within the production infrastructure are made available to controlled collaborations of users represented in the infrastructure through Virtual Organisations (VOs). Access to a VO is governed by a VO manager who is responsible for managing the addition and removal of users and the assignment of users to groups and roles within the VO. This capability is related to the ability of associating a user with a set of attributes in a trusted manner to a relying party, by way of digitally signed assertions.

#### 3.2.19.1 Available Standards

SAML from OASIS provides a standard interface for querying user attributes and also for the attributes format.

### 3.2.20 VirtualMachine.Management

The core functionality is for authorized users to manage the virtual machine life-cycle and configuration on a remote site (i.e. start, stop, pause, etc.). Machine images would be selected from a trusted repository at the site that would be configured according to site policy. Together this would allow site managers to determine both who could control the virtual machines running on their sites and who generated the images used on their site.

#### 3.2.20.1 Available Standards

The OCCI from OGF is an evolving specification useful in this area.

### 3.2.21 VirtualMachine.ImageFormat

This capability refers to the ability of porting virtual images on different hypervisor providers in order to increase the exploitation of the infrastructure.

#### 3.2.21.1 Available Standards

OVF (Open Virtualization Format) from DMTF simplifies interoperability, security, and virtual machine lifecycle management by describing an open, secure, portable, efficient, and extensible format for the packaging and distribution of one or more virtual appliances and applications. This enables software developers to ship pre-configured, ready-to-deploy solutions, allowing end-users to distribute applications into their environments with minimal effort.

### 3.2.22 VirtualMachine.ImageDistribution

As virtual machine images become the default approach to providing the environment for both jobs and services, increased effort is needed on building the trust model around the distribution of images. Resource providers will need a mechanism for images to be distributed, cached and trusted for execution on their sites.

#### 3.2.22.1 Available Standards

There are no available standards for this capability.

### 3.2.23 Operations.Monitoring

This capability relates to the capability of periodically observing measurements, transform them and make available to users or other applications. All of the resources within the infrastructure should be monitored for the community to be assured of the quality. Such a monitoring capability is essential for the operational staff attempting to deliver the production infrastructure and the end-users seeking out reliable resources to support their research.

#### 3.2.23.1 Available Standards

There are no available standards for this capability.

### 3.2.24 Operations.Accounting

This capability relates to the ability of systematically recording, reporting, and analyzing the usage of resources. The resource usage can be also used to perform billing depending on some agreed pricing model.

#### 3.2.24.1 Available Standards

The OGF defined both and interface (RUS) and a record format (UR) for accounting.

### 3.2.25 Client.API

A different approach to address interface heterogeneity is to build a client-side abstraction providing a common interface and taking care of translating the interactions into the specific implementation context.

#### 3.2.25.1 Available Standards

In this area, the OGF proposes the SAGA API.

# 4 REQUIREMENTS FOR STANDARDS EVOLUTION

This section should highlight the priorities for standardisation that EGI will pursue through its technology providers. Priorities will be identified together with the technology providers, Grid operators and application integrators and will be added here in future revisions of this document.

Even though we have not came across a formal collection of requirements for standards evolution, we can highlight three main priorities for the community that we anticipate will receive lots of attention:

- A standard job submission interface covering the functionalities of the main middlewares used in production is a long-standing issue in the Grid community; the PGI Working Group has been working in this area for the last two years; we expect that results of this group will fill the gap, enabling high-level service developers to rely on a single interface to manage job submission across different sites using different technologies
- A common information model across Grid middleware is also a missing gap; in this area, the GLUE 2.0 specification is being implemented by the various middlewares thus solving this issue in the next year
- Many educational institutions have a central identity management system used for federated authentication to external services based on the SAML 2.0 protocol; such institutions require the ability to bridge their identity systems to the Grid authentication system

# 5 CONCLUSION

Interoperability is a key requirement for EGI because participating systems and organizations require the ability to cross each other's boundaries in order to operate effectively together. It is widely recognised that open standards are key enablers for interoperability of networks, services and equipment.

With this document, we provided a survey of the most relevant standards for Distributed Computing Infrastructures highlighting the owning standard body and the adoption status. Further evolution of this work will focus on the identification of priorities together with the technology providers, Grid operators and application integrators.

# 6 REFERENCES


| | |
|---|---|
| **R1** | Definition of Interoperability - http://www.rand.org/pubs/monograph_reports/MR1235/MR1235.chap2.pdf |
| **R2** | EICTA White Paper on Standardisation and Interoperability - http://www.eicta.org/fileadmin/user_upload/document/document1166544474.pdf |
| **R3** | D5.1 UMD Roadmap - https://documents.egi.eu/document/100 |
| **R4** | W3C - http://www.w3.org/ |
| **R5** | WS-I - http://www.ws-i.org/ |
| **R6** | IETF – http://www.ietf.org/ |
| **R7** | OASIS - http://www.oasis-open.org/ |
| **R8** | OGF – http://ogf.org/ |
| **R9** | DMTF – http://www.dmtf.org/ |
| **R10** | Standards list - https://wiki.egi.eu/wiki/Standards |
| **R11** | IEEE - http://www.ieee.org/ |