# EGI-InSPIRE

## UMD ROADMAP

### EU DELIVERABLE: D5.1

| | |
|---|---|
| Document identifier: | EGI-D5.1-100-V01 |
| Date: | **04/11/2010** |
| Activity: | **SA2** |
| Lead Partner: | **EGI.eu** |
| Document Status: | **FINAL** |
| Dissemination Level: | **PUBLIC** |
| Document Link: | https://documents.egi.eu/document/100 |

Abstract

The Unified Middleware Distribution (UMD) is the integrated set of software components that EGI makes available from technology providers within the EGI Community. These components are packaged to provide an integrated offering for deployment on the EGI production infrastructure. This first version of the UMD Roadmap focuses on the overall structure of the UMD Roadmap in order to solicit feedback from the EGI Community. It presents a framework for technology providers to identify which capabilities they wish to address, and for the user and operations communities in EGI to identify additional capabilities they may wish to see supported within the infrastructure. The process by which this feedback will be integrated into future versions of the UMD Roadmap is described.

## I. COPYRIGHT NOTICE

## II. DELIVERY SLIP

|  | Name | Partner/Activity | Date |
|---|---|---|---|
| **From** | Steven Newhouse | EGI.eu | 23/7/10 |
| **Reviewed by** | **Moderator:** Jamie Shiers<br>**Reviewers:**<br>Johan Montgant, Tiziana Ferrari | SA2<br><br>SA2,EGI.eu | 26/7/10 |
| **Approved by** | **AMB & PMB** |  | 14/9/10 |

## III. DOCUMENT LOG

| Issue | Date | Comment | Author/Partner |
|---|---|---|---|
| 1.0 | 23/7/10 | First draft | Steven Newhouse, EGI.eu |
| 2.0 | 5/8/10 | Revised following reviews from JM and ML | Steven Newhouse, EGI.eu |
| 3.0 | 27/8/10 | Revised following comments from TF | Steven Newhouse, 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

This first version of the Unified Middleware Distribution (UMD) Roadmap introduces the framework that will be used in subsequent versions. It provides a framework partitioned into functional, security and operational capabilities that are important to different communities. This is a non-exhaustive list of capabilities that may be needed in a production infrastructure to support a research community. One of the purposes of this first version is to solicit feedback from the community on the priorities they attach to the identified capabilities and to identify additional areas for consideration.

The UMD Roadmap will be formally reviewed and re-issued every six months during the course of the project, but will also exist as a 'living' document in between these releases to allow a continual process of community comment and feedback. The process by which input from the EGI Community is collected prioritised and integrated into the UMD Roadmap through the Technical Coordination Board (TCB) is also described.

**TABLE OF CONTENTS**

# 1 INTRODUCTION

## 1.1 *Purpose*

This document describes the framework that will be used to build the Unified Middleware Distribution (UMD) Roadmap and the process by which contributions from the community will be integrated into future versions of the UMD Roadmap. The UMD Roadmap therefore describes the structure of the UMD and the process by which its functional capabilities will evolve over time in response to feedback from the end-user and operations communities. It is a document that will be formally reviewed and re-issued every six months during the course of the project, but will exist as a 'living' document in between these times to allow a continual process of community comment and feedback. This first version of the UMD Roadmap provides a framework for the technology providers and the community as a whole to comment upon.

## 1.2 *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.

## 2 TECHNOLOGY WITHIN EGI

EGI provides a production quality e-Infrastructure to researchers across Europe which interoperates with other e-Infrastructures worldwide. This activity is driven by the EGI user communities that are shared between such infrastructures. This document introduces a number of important concepts in defining the EGI Technology and UMD Roadmaps. These relate to:

- A *technology provider* is an organisation or a project that is collaborating with EGI to develop or deliver software for use within the production infrastructure or its user community.

- A *software component* is provided by a technology provider in a software release and is the smallest granularity of software considered within this document.

- The *Unified Middleware Distribution* (UMD) is the set of software components developed for EGI by technology providers to provide the innovation needed by EGI to satisfy its users that cannot be found elsewhere and is therefore endorsed by EGI for use within the production infrastructure.

- An EGI *Capability* (e.g. authorization, job submission, data movement, etc.) is a high-level description of activity needed by either the end-user or operations community that is defined and delivered by one or more I*nterface*s that may be supported by one or more technology providers.

- An EGI *Interface* consists of different functional aspects that together or separately are verified through the component acceptance criteria, i.e. different functional aspects of the interface have defined acceptance criteria and are verified by the EGI.eu Technology Unit. These interfaces may initially be specified through free text but will over time be defined by a particular protocol, specification or standard.

- An EGI *Functionality* is provided as part of an EGI Interface to meet a requirement identified by the end-user or operations communities within EGI.

- The UMD *Release* is the set of components within UMD that are released to provide an integrated software distribution. There may be components in UMD that are not included in the UMD Release but provide functionality that is only available from the EGI Software Repository.

- All software components, interfaces and releases are defined by a *version numbering scheme* that follows the conventional major.minor.patch (e.g. 3.1.4) format. A patch release must preserve the same interfaces (CLI and API) and protocol/specification versions. If any of these interfaces or protocol/specification is extended, but retains compatibility with earlier releases, then that difference is recognized by increasing the minor release number. If a release is not backwards compatible then the major release number is increased.

- The EGI *Platform* is the collection of UMD Interfaces that is implemented within UMD. Note – this is distinct to the underlying OS environments supported within EGI.

- The *EGI Community* are the 'DCI projects', other national or European Commission funded projects, the European National Grid Initiatives and the user communities benefiting from the European Grid Infrastructure.

- The *EGI Technology Roadmap* provides a broader view to the use of technology within the EGI production infrastructure, e.g. supported operating systems, messaging technologies, deployed software, etc. The EGI Technology Roadmap includes components sourced from outside the EGI Community.

---

- The *UMD Roadmap* defines the evolution of UMD, the software components coming from within the EGI Community, which will be deployed to deliver the production infrastructure.
- The *Technology Coordination Board* (TCB) brings together stakeholders within the EGI Community to prioritise their requirements into the broader EGI Technology and the UMD Roadmaps which will be issued to together at least every 6 months.

## 2.1   EGI Technology Roadmap

EGI has to provide a technical infrastructure that needs to satisfy many stakeholders. Its most important class of stakeholder are the end-user communities that span many diverse disciplines, physically distributed across the European Research Area. These communities have many different approaches to use the infrastructure and different technologies are needed for federating the distributed resources provided by the collaborating resource providers in order to support their end-user communities. The resource providers are generally federated into national or domain specific groups (i.e. NGIs or EIROs) and bring together the individual sites from a resource provider into an integrated operational and management structure. The long-term evolution of the production environment delivered by the resource providers is defined within the EGI Technology Roadmap where, over time in order to achieve sustainability, a greater proportion of the technology should ideally be coming from open-source and commercial providers outside of the EGI Community.

The EGI Technology Roadmap, which includes the UMD Roadmap, will over time aim to:
- Define technologies that will help EGI resource providers to reliably deliver services to more communities with smaller effort than they do today.
- Define stable platform interfaces that allow the user and developer communities to add value through their own software into the EGI environment.
- Define a set of software components from within (i.e. the UMD Roadmap) and external to the EGI Community drawn from open-source or commercial technology providers that are able to contribute to delivering the production infrastructure, some of which may be made available through the EGI Software Repository
- Identify a programme of support, maintenance and development for the software components that form the core functionality within the production infrastructure built on top of the base operating systems.

Through these aspects, the EGI Technology Roadmap defines the technical architecture of the software deployed on the European Grid Infrastructure and the environment in which the Unified Middleware Distribution (UMD) has to operate. For instance, the EGI Technology Roadmap would specify software that is drawn from outside the community such as the base operating system (e.g. Scientific Linux or Ubuntu) or the use of a particular messaging system (e.g. ActiveMQ). The majority of these components will initially be sourced from within the community and integrated into a UMD which is described within the UMD Roadmap. Other components from commercial providers or components from outside the EGI Community are described within the EGI Technology Roadmap. Ultimately, the technology used to deliver the production environment will, in order for it to be sustainable, need to come from providers both within and external to the EGI Community. The EGI Technology Roadmap will be issued alongside the UMD Roadmap by the TCB. The components within both the UMD and the EGI Technology Roadmaps are described in terms of their *capabilities* (i.e. consisting of specific *functionality* aspects that will be delivered in response to particular user

communities over time) delivered through the production of *software releases* from a set of *technology providers*. These technology providers may deliver open-source software from within or external to the EGI Community (but outside the EGI-InSPIRE project) or relevant commercial software solutions. This allows the EGI to be technology neutral and to respond to the needs of its users – both the researchers using the infrastructure and the operations staff providing it – by selecting the best available technology for delivering the capabilities that are required.

## 2.2   UMD Roadmap

UMD is the set of software components that the EGI Community needs to provide to fill the gap between the functionality it needs to deliver as a generic base to its user community and the 'out of the box' solutions available from mainstream commercial or open-source providers. The UMD Roadmap describes the capabilities of the software within UMD and how the functionality within each capability will evolve over time in response to the requirements coming in from the community. The current software used to deliver the production infrastructure to end-users can be grouped into three areas:

- Components that provide functionality required by a particular community and are best supported by teams within that community. They are deployed into the environment provided by the EGI Technology Roadmap.

- Components from outside the EGI Community described in the EGI Technology Roadmap. These may be components that have emerged from within the EGI Community and have now reached sufficient maturity, stability and functionality such that they have use beyond the e-infrastructure community. One route towards broader adoption is to contribute these components to mainstream software distributions and to use their associated communities as a way to expand the support structures. These components are then imported back into the EGI Community through the EGI Technology Roadmap defining these software distributions as a basis for the infrastructure. The EGI Technology Roadmap could also include commercially provided/supported software.

- Components that are still being developed or deliver functionality that is only needed within the e-Infrastructure community as described in the UMD Roadmap.

Over time it is expected that components will move between these three categories as they mature and their uptake and usage changes. Components in the second group, i.e. components contributed to mainstream open-source software distributions, will help reduce the bespoke integration needed within EGI and promote the uptake of these components by a wider (non-Grid) community. Such activity contributes greatly to the sustainability of these components as they are exposed to a much wider user base than just a deployment within UMD. The availability of such capabilities within these distributions and their expected functionality will be described in the EGI Technology Roadmap. As the user communities mature, many of these capabilities are expected to be satisfied by the offering provided by the services within EGI. Those communities requiring additional specialised capabilities beyond those described in the UMD Roadmap and EGI Technology Roadmap will need to demonstrate that their capability has applicability to other user communities, or to develop and deploy these capabilities themselves.

How the number of components in each group
could vary over time
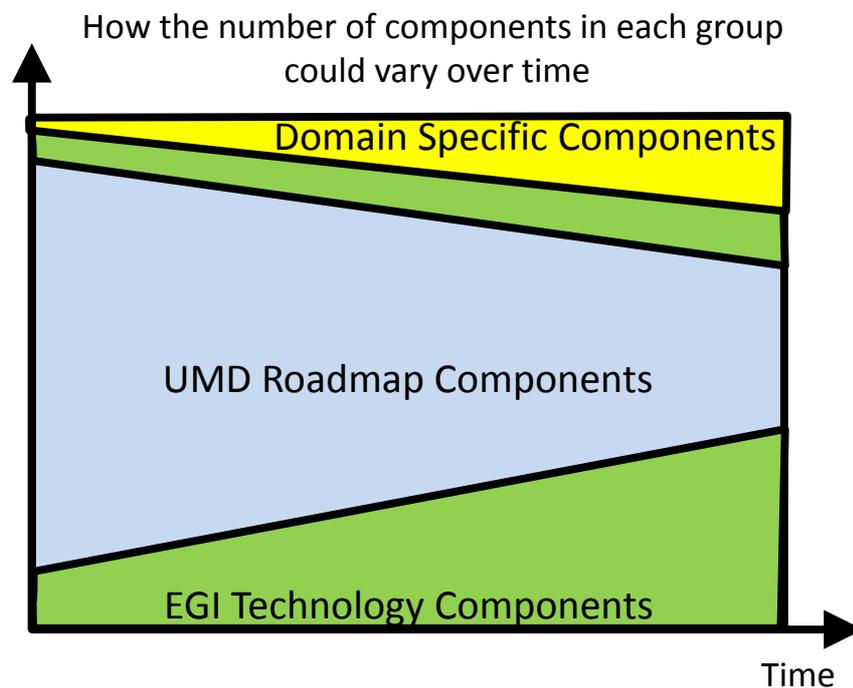


The UMD Roadmap describes the technical architecture of the UMD and documents how it is expected to change over time – the components that will be added or removed and the functionality that will appear or be deprecated. The roadmap will effectively be a living document – requirements will change or be refined over time and the technology providers will adapt their release schedules. Formally, the changes in the UMD Roadmap will be reviewed every six months with changes being made available between these formal reviews to solicit community feedback.

## 2.3 Governance

The governance of both the broader EGI Technology Roadmap and the UMD Roadmap, in terms of the prioritisation of new requirements, the inclusion of new technology providers and new components is driven by the Technology Coordination Board (TCB) [R4] run by EGI.eu on behalf of the EGI community. The composition of the TCB is made up of voting members:

- Chief Technology Officer (Chair)
- Chief Operations Officer representing the Operations community.
- Chief Community Officer representing the end User community.
- Technology providers that are committed to contributing software into UMD.
- Deployed Middleware Support Unit team leader (TSA2.5).

and observers from the other SA2 activities are present to provide input relating to the:

- definition of quality criteria (TSA2.2).
- verification of quality criteria (TSA2.3).
- EGI Software Repository (TSA2.4).

Representatives of the Technology Providers that have committed to deliver software components to EGI.eu for inclusion in the UMD – generally expressed by a Memorandum of Understanding (MoU) and/or Service Level Agreement (SLA) – will be part of the TCB.

## 2.4 Virtuous Cycle
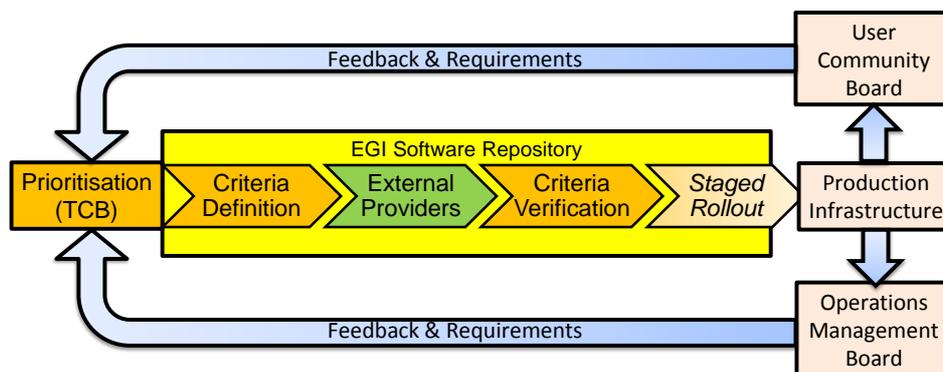


The virtuous cycle established in the EGEE series of projects is adopted to define the EGI Technology Roadmap which describes the evolution of the production infrastructure. Within the EGI Technology Roadmap is the UMD Roadmap which further describes the software components developed within the EGI Community which are not available from other sources and therefore have to be supported by the EGI Community. For both Roadmaps the prioritisation of the requirements coming from the consumers of the technology – the operations and the end-user communities – takes place within the TCB where technology providers are represented alongside the CCO and COO respectively representing the end-user and operations communities.

Once these requirements have been prioritised the work of sourcing these capabilities can begin. The strategy for fulfilling these requirements will depend on the identified functionality. The TCB will rely on the knowledge from within the TCB and suggestions collected from the community. Adaptation of an existing similar component provides an incremental approach. Working with technology providers within the EGI Community to go through a detailed design, prototyping, development and deployment model is another. Collaboration with a commercial technology provider to deliver the capability could also be explored. By publishing these requirements and widely promoting these needs, contributors can be expected from outside the immediate EGI community in Europe and attract software from other technology providers around the world.

The components proposed to meet the published requirements will be assessed by EGI.eu Technology Unit and by soliciting feedback on the effectiveness of the software from the end-user and operations communities. The results from these assessments will be published and considered by the TCB in making their decision as to how the particular technology solution can be best integrated into EGI.

It is the TCB that manages the relationship that EGI has with the 'Technology Providers' that it is using to deliver its software infrastructure. The software is delivered by these providers according to 'criteria' (both generic and specific to particular functionalities) that have been derived from the requirements needed by the operations and user communities. The delivered software components

are verified against these criteria to ensure they meet the specified requirements. The 'Verification' process used can be adjusted dependent on the scope of the release and the quality assurance processes used by the technology provider. Once verified the software moves into 'Staged Rollout' where the new release is deployed to early adopter sites, before a wider scale deployment in the infrastructure. The staged rollout allows for any blocking flaws in the new release to be detected and corrected before a large number of sites and/or users might be affected.

Once the technology is deployed and in use, feedback and new requirements are collected and prioritised locally by the two key stakeholders:

- the NGI Operations Managers (representing the resource providers) through the Operations Management Board
- the User Community through the User Community Board

## 2.5 Roadmap Structure

The EGI Technology Roadmap, and within it the UMD Roadmap, are both structured around *capabilities*. These capabilities are classified into three areas:

- 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.
- Managerial: A managerial capability is provided to help with the operation of the functional capability within the production infrastructure.

An example of such a structure is presented below:

**End-User Community**

| | | |
|---|---|---|
| Security | **Authentication Capability** | |
| | **Authorisation Capability** | |
| Functional | **Computing Capability** / **Data Capability** / ---- / ---- / **Information Discovery Capability** | |
| | **Monitoring Capability** | |
| Operational | **… Capability** | |
| | **Management Capability** | |

**Operations Community**
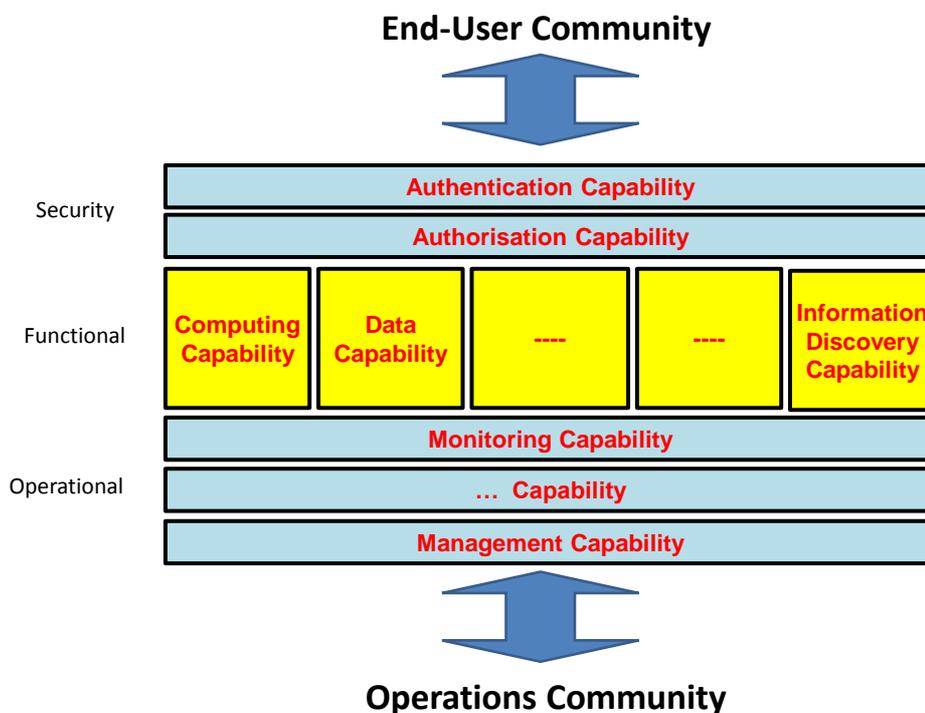
This figure identifies the need for consistency across all the functional capabilities (that the operations and end-user communities are looking to introduce into the infrastructure) for security and operation of the infrastructure. Consistency in security is essential - so there is a common approach to authentication and authorisation across all functional capabilities that are exposed from within a site. In order to reduce the management and operational burdens for the resource providers, implementations of the different capabilities should have consistent approaches to configuration, monitoring, accounting, etc. A consistent approach to service administration and management that is aligned with existing practices will simplify the adoption of grid technology within new communities and reduce the burden to existing communities.

Each capability supports one or more interfaces. An interface within a capability delivers one or more functionalities. The functionality that makes up an interface is implemented by one or more software components. These software components may be provided by one or more technology providers.

## 2.6  Support Levels

Components within the EGI Technology Roadmap and the UMD Roadmap will have identified support levels. The EGI Technology Roadmap will focus on components coming from outside the community with broader and probably more established support structures. The UMD Roadmap, with components coming from within the EGI Community, will have varying levels of support driven by the areas of the community they are drawn from: These support levels are described below:

- Integrated: Components for which the technology provider asserts that they have met the specified criteria and this has been verified by EGI.eu and accepted by the TCB for integration into UMD. The component may contribute towards a reference implementation of the UMD Platform. These components will be covered under a SLA signed between EGI.eu and the relevant technology provider. The software components are able to be deployed alongside each other on a single host – although operational deployments may spread services out for load balancing and other reasons.

- Contributed: Components for which the provider asserts that they have met the specified criteria and this has been verified by EGI.eu and that have been accepted by the TCB for distribution (not integration) with UMD. This provides a mechanism to distribute components that are useful within UMD without EGI having to fully integrate them into UMD. Depending on the adoption of such a component by the community the TCB may decide at a later date to 'integrate' this component. The support offered by the technology provider may not be covered by an SLA.

- Candidate: Components for which the provider and EGI have agreed to work together to improve the current quality and the software has not yet reached the quality to be endorsed by the TCB. However, the TCB believes there is sufficient interest from the EGI Community to distribute the component with UMD and to promote its use.

- Offered: Components for which the provider feels that the software has demonstrated some useful functionality which has been recognised and requested by part of the EGI community. The maintenance and quality of the software may not be ensured at this time by EGI.eu, but rather directly by the technology provider. This provides a mechanism for new components being developed within the community to gain broader exposure within the EGI community by inclusion into the EGI Software Repository and their availability being disseminated more widely.

## 2.7 Process

### 2.7.1 Functional and non-functional Requirements

The functional and non-functional requirements from the end-user and operations communities are collected initially through the User Community Board (UCB) and the Operations Management Board (OMB) respectively. It is expected that some filtering and prioritisation of requirements will take place within these bodies. Functional requirements relate to the behaviour of the software component – the outputs it provides in response to the provided input. Non-functional requirements relate to the quality of the of the software component – its response time, throughput, uptime, etc.

The TCB provides an opportunity to identify multiple requirements from these different communities that can be collected into a single capability or added to an existing capability. These capabilities are versioned as a decision could be made to add a new functionality into an existing capability, but in order to allow time for implementers to support this capability the new version would only be introduced at a specified date.

### 2.7.2 Technology Providers

Technology providers wishing to have their components considered for inclusion into the UMD Roadmap should email cto@egi.eu. They should indicate:

- The capability in the UMD Roadmap (identified in sections 4, 5 & 6 of this and future versions of this document) addressed by the component.
- The current user community for the software.
- The current support structure for the software.
- The desired level of support (integrated/contributed/candidate/offered).

At regular intervals the TCB will consider such requests drawing on the expertise and knowledge within the TCB, the DMSU, and the broader EGI Community as to the suitability of the component. If adopted into the UMD Roadmap usage of the component will be monitored and the support level revised as appropriate.

All technology providers wishing to achieve integrated, contributed or candidate status will be expected to enter into a Memorandum of Understanding (or an equivalent contractual commitment for a commercial provider) with EGI.eu as to the support commitment they are willing to provide. Those technology providers wishing to achieve integrated status will be expected to enter into a Service Level Agreement (SLA) or equivalent with EGI.eu to provide support for their components and be integrated into EGI.eu's support structure.

It is possible for the TCB to endorse more than one component that answers the need for a particular capability in the UMD Roadmap. This allows different technical approaches to delivering the same capability to be explored and different implementations tuned to the different deployment environments that need to be provided to satisfy the EGI community's diverse needs. The ability for the TCB to endorse multiple implementations of the same capability is dependent on the definition and adoption of common interfaces by the different technology providers. This is an essential pre-condition in order that end-user communities are able to transparently use either deployed implementation without having to adapt their client code or workflow.

# 3 FUNCTIONAL CAPABILITY

A functional capability is defined by a set of interfaces that meets the particular needs of a community. These interfaces are defined by a set of criteria – both functional and non-functional – that need to be met by the technology provider. These criteria will be split into two areas:

- generic criteria that are applied across all capabilities
- specific criteria that are applied to a particular capability

It is noted that implementations for many of these capabilities are already in production in infrastructures provided through projects such as DEISA, EGEE and NorduGrid in Europe, and projects such as Open Science Grid and TeraGrid in the US. No attempt has been made to categorise implementations against these capabilities in this version of the UMD Roadmap in order to leave it open to the technology providers as to which capabilities they wish to deliver solutions to.

## 3.1 Information Discovery Capability

### 3.1.1 General Description

Information discovery is a service that helps find the required resources that have been registered with it within the production infrastructure. Resources within the infrastructure need to be described using a consistent and extensible schema that is adopted by all technology providers. The collected information is made available through well-known instances that provide data relating to some logical collection, infrastructure wide, regional, site, domain, etc.

### 3.1.2 Supported Interfaces

The LDAPv3 (RFC 4530) protocol and search syntax is used to query information from the information discovery services and to encapsulate the information payload relating to the services being offered within the production infrastructure that is exchanged between instances. The information about the resources is described using the GLUE schema from the Open Grid Forum. Currently this is GLUE 1.3 with migration underway to GLUE 2.0 [R5].

### 3.1.3 Implementation Roadmap

Implementations expected from EMI and IGE.

## 3.2 Compute Capability

### 3.2.1 General Description

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.

### 3.2.2 Supported Interfaces

There are a number of different proprietary interfaces currently in production use that provide the ability to describe, submit and manage an interactive or batch work item on a specific site. Activity within the Open Grid Forum 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 and the experiences derived from them 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.

### 3.2.3  Implementation Roadmap

Implementations are expected from EMI and IGE. It is expected over time that these will converge on a defined interface.

## 3.3  *Compute Job Scheduling Capability*

### 3.3.1  General Description

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.

### 3.3.2  Supported Interfaces

The implementations provided for Compute Job Scheduling should use compatible interfaces for the batch compute capability.

### 3.3.3  Implementation Roadmap

Implementations expected from EMI, IGE and other job scheduling services.

## 3.4  *File Access Capability*

### 3.4.1  General Description

Provides an abstraction that allows a file to be stored on or retrieved from a storage device (e.g. tape, disk, distributed file system, etc.) for use elsewhere in the infrastructure.

### 3.4.2  Supported Interfaces

The Storage Resource Management (SRM) interface specification [R3] is a commonly adopted specification in this area which defines much of the functionality used in this area. 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.

### 3.4.3  Implementation Roadmap

Many implementations exist in the community.

## 3.5  *File Encryption/Decryption Capability*

### 3.5.1  General Description

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.

### 3.5.2 Supported Interfaces

The key handling interface will be described in future versions of the roadmap following input from the EGI Community.

### 3.5.3 Implementation Roadmap

Implementation expected from EMI.

## 3.6 Database Access Capability

### 3.6.1 General Description

Many communities are moving to the use of structured data stored in relational databases. These need to be accessible for controlled use by remote users as any other e-Infrastructure resource.

### 3.6.2 Supported Interfaces

To be described in future versions of the roadmap following input from the EGI Community.

### 3.6.3 Implementation Roadmap

This is an area where implementations are maturing.

## 3.7 Metadata Catalogue Capability

### 3.7.1 General Description

The metadata catalogue is used to store and query information relating to the data (files, databases, etc.) stored within the production infrastructure.

### 3.7.2 Supported Interfaces

To be described in detail in future versions of the roadmap following input from the EGI Community. 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.

### 3.7.3 Implementation Roadmap

Implementations exist within EMI and IGE.

## 3.8 File Transfer Capability

### 3.8.1 General Description

Files are stored at different physical locations within the production infrastructure and are frequently used at other locations. It is necessary for the files to be efficiently transferred over the international wide area networks linking the different resource centres.

### 3.8.2 Supported Interfaces

The GridFTP protocol [R2] is used extensively in production infrastructures around the world alongside protocols such as http/https that have been developed outside of this community. This protocol provides the functionality to read/write and list data files stored on remote locations.

### 3.8.3 Implementation Roadmap

There are many implementations supporting various file transfer protocols.

## 3.9 File Transfer Scheduling Capability

### 3.9.1 General Description

The bandwidth linking resource sites is a resource that needs to be managed in the same way compute resources at a site are accessed through a job scheduler. By being able to schedule wide area data transfers, requests can be prioritised and managed. This would include the capability to monitor and restart transfers as required.

### 3.9.2 Supported Interfaces

To be described in future versions of the roadmap following input from the EGI Community.

### 3.9.3 Implementation Roadmap

Implementations are expected from EMI and IGE.

## 3.10 Parallel Job Capability

### 3.10.1 General Description

The parallel programming paradigm is gaining greater use in the user communities within EGI. The infrastructure does not provide support at the programming level – it is not needed – 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.

### 3.10.2 Supported Interfaces

There are several parallel programming paradigms that need to be supported. These include:

- MPI: Message Passing Interface 1.x [R6]
- MPI: Message Passing Interface 2.x [R7]
- OpenMP [R8]

### 3.10.3 Implementation Roadmap

Contributions are expected from EMI and IGE.

## 3.11 Remote Instrumentation Capability

### 3.11.1 General Description

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.

### 3.11.2 Supported Interfaces

To be described in future versions of the roadmap following input from the EGI Community.

### 3.11.3 Implementation Roadmap

See examples from the GridCC, DORII and DILIGENT projects.

## *3.12 Workflow Capability*

### 3.12.1 General Description

The ability to define, initiate, manage and monitor a workflow is a key capability across many user communities. It is also a capability that can be deployed by a user or a user community (i.e. it does not need to be a service provided as part of the core infrastructure) but the various workflow systems may have requirements that need to be supported within the core infrastructure.

### 3.12.2 Supported Interfaces

To be described in future versions of the roadmap following input from the EGI Community.

### 3.12.3 Implementation Roadmap

There are many workflow engines and it is not the purpose of this capability to implement workflow engines but to ensure that the EGI Platform provides the interfaces needed to support their use.

# 4 SECURITY CAPABILITIES

## 4.1 Authentication Capability

### 4.1.1 General Description

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.

### 4.1.2 Supported Interfaces

The primary authentication token within the infrastructure is the X.509 certificate and its proxy derivatives. The certificates and any proxy schemes must follow specifications that are fully integrated into the https protocol (as opposed to the httpg protocol). Wherever possible the user should not be exposed to certificates and their associated handling.

### 4.1.3 Implementation Roadmap

This capability is widely used in many middleware distributions.

## 4.2 Credential Management CAPABILITY

### 4.2.1 General Description

The Credential Management capability provides an interface for obtaining, delegating and renewing authentication credentials by a client using a remote service.

### 4.2.2 Supported Interfaces

One of the key functionalities in this area is the linking of institutional authentication systems (e.g. Kerberos, Shibboleth) to the transparent issuing of certificates for use in the infrastructure through identity federations. This should be provided for operational deployment through the use of web portals and web service interfaces.

### 4.2.3 Implementation Roadmap

See for example the work on Short-Lived Certificate Services (SLCS) and the Terena Certificate Service (TCS).

## 4.3 User Management Capability

### 4.3.1 General Description

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.

### 4.3.2 Supported Interfaces

An interface is needed to issue proxy certificates relating to the roles and groups that an individual has within a VO. The corresponding attributes may also be delivered back to the client through SAML assertions.

### 4.3.3 Implementation Roadmap

Contributions are expected from EMI and IGE.

## *4.4 Authorization Capability*

### 4.4.1 General Description

The implementation of access control policy – authorisation – needs to take place on many levels. Sites will wish to restrict access to particular VOs and individuals. Sites or VOs may wish to stop certain users accessing particular services. The infrastructure as a whole may need to ban particular users from the whole infrastructure. Policy Enforcement Points (PEPs) will be embedded into many components throughout the infrastructure and will use Policy Decision Points (PDPs) to drive access control decisions.

### 4.4.2 Supported Interfaces

APIs and libraries are needed for the integration of PEPs into software and a defined interface for the PDP is needed to allow different PDPs to be deployed within the infrastructure.

### 4.4.3 Implementation Roadmap

Implementations are expected from EMI and IGE.

# 5 OPERATIONAL CAPABILITY

## 5.1 Virtual Image Management Capability

### 5.1.1 General Description
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.

### 5.1.2 Supported Interfaces
To be described in future versions of the roadmap following input from the EGI Community.

### 5.1.3 Implementation Roadmap
To be determined following community input.

## 5.2 Virtual Machine Management Capability

### 5.2.1 General Description
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.

### 5.2.2 Supported Interfaces
To be described in future versions of the roadmap following input from the EGI Community.

### 5.2.3 Implementation Roadmap
To be determined following community input.

## 5.3 Messaging Capability

### 5.3.1 General Description
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.

### 5.3.2 Supported Interfaces
To be described in future versions of the roadmap following input from the EGI Community.

### 5.3.3 Implementation Roadmap
This capability can probably be achieved through the configuration of an existing open-source technology.

### 5.4 Monitoring Capability

#### 5.4.1 General Description

A production infrastructure is defined by its availability and reliability – if its user community cannot rely on it then it is not an infrastructure. All of the resources within the infrastructure need to 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.

#### 5.4.2 Supported Interfaces

Both programmatic and web based interfaces are needed to access this capability.

#### 5.4.3 Implementation Roadmap

The monitoring capability is implemented through probes provided for each functional capability that is being monitored and a framework within which these probes reside. The probes are provided by either the operations teams or the technology providers, and are integrated into a framework deployed within the production infrastructure.

### 5.5 Accounting Capability

#### 5.5.1 General Description

The use of resources within the e-Infrastructure must be recorded for understanding usage patterns by different user communities and by individuals within their communities.

#### 5.5.2 Supported Interfaces

A consistent interface and message format is required for different resource types to record the use made of their capabilities by different users and to inject these records into the accounting store.

#### 5.5.3 Implementation Roadmap

Implementation is needed of an accounting store, which may also be deployed on a national or regional basis, with a central instance so that consistent metrics relating to the use of the infrastructure can be derived. The generation of the accounting records is the responsibility of the technology provider.

# 6 EGI PLATFORM

The EGI Platform is the collection of EGI Interfaces from the different EGI Capabilities and represents a commitment between the technology provider and the consumer of the interface (end-user, developer or operations) for stability and planned evolution that will be managed by the TCB and validated by the EGI.eu Technology Unit on behalf of the community. The EGI Platform is expected to evolve into a stable set of service interfaces or programming APIs (depending on the interface) that can be considered a solid foundation upon which to build services, applications or tools. The EGI Platform as a whole, and the individual Interfaces are versioned and changes are planned as part of a software component's release roadmap. Any changes to an interface are planned in consultation with its users and announced as part of the roadmap.

Wherever possible the EGI Interfaces will reference defined standards interfaces (from organisations such as the Open Grid Forum, IETF, W3C, OASIS, etc.), protocols or APIs provided by third parties. Advice is given to the community about interfaces that are emerging from the community and are 'candidate' interfaces within the UMD Roadmap. These interfaces are potentially subject to rapid change. If there is no consensus in the community on a common interface for a particular capability then such gaps are also noted.

The functional and non-functional aspects of the interfaces are defined and verified by the EGI.eu Technology Unit through quality criteria. Early drafts of these quality criteria for some capabilities are available [R1] and these will be expanded during the course of the project.

# 7 CONCLUSION

This document describes how the technological evolution of the production infrastructure will be planned within EGI. Key to this planning is the source of software components coming from technology providers outside the infrastructure that will deliver the technological innovation to meet the evolving needs of the end-user and operations communities. The software components fulfilling these requirements coming from within the EGI community will be described in the UMD Roadmap.

The first release of the UMD Roadmap (August 2010) describes the framework that will be used to structure the requirements of the end-user and operations communities, and the software components available from the technology providers. Input is expected from the community in identifying:

- Additional capabilities not currently listed
- Prioritisation of the listed capabilities
- Key interfaces that need to be supported to access these capabilities
- Implementations from technology providers able to deliver these capabilities

Feedback on the UMD Roadmap can be sent to cto@egi.eu

# 8   REFERENCES

| R 1 | https://wiki.egi.eu/wiki/EGI-InSPIRE:UMDQualityCriteria |
|-----|----------------------------------------------------------|
| R 2 | http://ogf.org/documents/GFD.20.pdf |
| R 3 | http://www.ggf.org/documents/GFD.129.pdf |
| R 4 | https://documents.egi.eu/document/109 |
| R 5 | http://www.ggf.org/documents/GFD.147.pdf |