Deliverable D3.2

Standardization Roadmap and First Sustainability Plan

Project acronym: MAPPER

Project full title: Multiscale Applications on European e-Infrastructures.

Grant agreement no.: 261507

Due-Date:	30 September 2011
Delivery:	28 September 2011
Lead Partner:	LMU
Dissemination Level:	Public
Status:	Final
Approved:	
Version:	1.0

DOCUMENT INFO

06.07.2011	W. Dubitzky	initial version
18.08.2011	M. Schiffers	Methodology, resources, roadmap
14.09.2011	W. Dubitzky	Science, ICT R&D added. Summary, Introduction, Conclusion
17.09.2011	M. Schiffers	Finalized version 1.0 for internal review
19.09.2011	A. Hoekstra	Internal reviewing
23.09.2011	C. Kurowski	Software chapter and adaptations
23.09.2011	M. Schiffers	Integration of reviewer’s comments

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Executive Summary

The MAPPER project develops concepts, methods and technology facilitating the development, deployment and execution of multiscale modelling and simulation applications. To ensure that the technology developed by MAPPER is adopted and further developed, the project partners are specifying a MAPPER Standardization Roadmap and a MAPPER Sustainability Plan. This document contains the initial version of both.

The sustainability approach is structured into the following four layers of tasks and activities: Science, ICT R&D, Software, and Resources. Each layer is intended to be associated with various standardization and sustainability tasks and activities.

This document also defines an initial version of the MAPPER Standardization Roadmap. The Roadmap describes the standards relevant to ongoing MAPPER implementations and mid-term to long-term issues to be solved for e-infrastructures capable of Distributed Multiscale Computing.

The document is a living document which will be updated in month 24 by deliverable D3.6.

1 Introduction

The declared goal of MAPPER is to ease the development and deployment of Distributed Multiscale Computing (DMC) for multiscale *modelling and simulation* (MMS) applications in emerging and future e-infrastructures. With this goal in mind and based on a selection of MMS applications from various communities, the MAPPER project is developing novel DMC concepts, methods and tools. Key to the success of the MAPPER project is the sustainability of the developed technology beyond the life time of the project. The MAPPER sustainability approach is characterized by the dimensions depicted in Figure 1.

On one hand, MAPPER is concerned with tasks, activities and procedures: *operation/deployment*; *services*; *dissemination, training* and *support*; and *policies*, *standards* and *interoperability*. These dimensions are associated with different aspects of science and engineering, both in computer science (*ICT R&D*, *software*, *resources*) and in scientific areas that deploy and use MMS applications (*science*). Figure 6 provides a more elaborate depiction of these dimensions and their interrelationships.

	Operation/Deployment	Services	Dissemination, Training, Support	Policies, Standards, Interoperability
Science				
ICT R&D				
Software				
Resources				

Figure 1: Dimensions of the sustainability plan.

The basic rationale behind the four layers is as follows:

1. **Layer I: Science** is focusing on the scientific communities in the various areas where MMS is increasingly used to tackle complex scientific challenges. This layer is concerned with supporting these communities in the deployment and use of MAPPER DMC tools. Additionally, MAPPER envisages developing new teaching curricula from an e-science (with emphasis on DMC) perspective.
2. **Layer II: ICT R&D** revolves around computing concepts, methods, data and information structures/models and algorithms developed by MAPPER. Sustainability requires publishing them in suitable scientific journals, at workshops and conferences, and at the MAPPER seasonal schools with the objective to facilitate the use and further development of MAPPER methodologies and technologies. A detailed discussion on the MAPPER sustainability ICT R&D efforts and activities is presented in Section 5.
3. **Layer III: Software** forms one of the key technology contributions of the MAPPER project and is central to MAPPER's approach to sustainability. We distinguish two levels of software (see Section 4): *high-level software* and *low-level software*. While low-level software relates to basic programming and infrastructure services, high-level software is closely associated to the tools and components provided by *virtual research communities* (VRC). MAPPER partners and VRC members will adapt, maintain, further develop, and publicize both low-level and high-level software as far as they relate to MAPPER. The MAPPER Seasonal Schools, the MAPPER dissemination deliverables, and the MAPPER standardization efforts will additionally contribute to sustaining software by educating users and by contributing to standardization and policy bodies. A precise software sustainability plan depends on the final list of software modules. For the initial version of this report, such a list is rudimentary. The unavoidable gaps will be closed in a final version of this document to be provided at the project's end.
4. **Layer IV: Resources** (e.g., network, computing and storage elements) are critical to deploying and running large-scale DMC applications. In the MAPPER context, access to such resources by VRCs will be via *virtual organizations* (VO). The MAPPER sustainability approach considers resources from four different perspectives: operational, service, educational and interoperability (Section 3 contains more on this.).

The MAPPER Sustainability Plan is accomplished by the Standardization Roadmap (Appendix A), which describes the standards relevant to ongoing MAPPER implementations and mid-term to long-term standardization issues in relation to emerging and future e-infrastructures capable of running DMC applications. The Roadmap is complemented by future MAPPER deliverables like the MAPPER Profile (Deliverable D3.3), the MAPPER Test Suite (Deliverable D3.5), and the Report on the Policy Framework Resource Providers Need to Adopt to Support the MAPPER Project (Deliverable D3.1).

Finally, to facilitate the future maintenance and use of MAPPER results and technology in the context of the wider development of DMC e-infrastructures, the MAPPER sustainability efforts will be complemented by a Foresight Study into New Concepts and Technologies for Multiscale Modelling on Large E-Infrastructures (Deliverable D3.4).

The remainder of this document is organized as follows: Section 2 outlines the methodology we have adopted to develop the MAPPER Standardization Roadmap and Sustainability Plan. Sections 3-6 describe the main areas of sustainability across the four layers (as relevant for the initial version of this report). Section 7 provides an action plan, Section 8 concludes the report.

2 Methodology

Both the Standardization Roadmap and the Sustainability Plan have been derived from several sources. While the Standardization Roadmap takes into account comparable efforts in similar projects and current work of standardization bodies, the Sustainability Plan discusses the sustainable availability of current (and planned) MAPPER resources and services.

Both plans are intertwined: The Roadmap aims at supporting sustainability and the Sustainability Planning contributes to the definition of the Standardization Roadmap. Consequently, the purpose of this deliverable is to outline the standards influencing the MAPPER project and the standards influenced by the MAPPER project. The goal is to derive a Roadmap for externalizing such standards for adequate compliance checking.

Although this document presents a Standardization Roadmap and a Sustainability Plan, the reader should be aware that the document can only reflect the current state of affairs. Bearing in mind that the standardization landscape is in a constant state of flux, this document needs to be considered a living document. This initial version sets the scene for future activities in this domain; an updated version will be available as deliverable D3.6 (month 24).

2.1 Introductory Note

Standardization is essential for the uptake and future development of services as envisioned by MAPPER in Europe and beyond. However, while standardization is a tedious and time-consuming task, reality has shown that projects often require faster turn-around times when producing their applications and underlying services. As a result, services developed in different projects may be implemented differently – without standardization – which typically results in a set of non-interoperable systems.

As the MAPPER project aims at leveraging European e-infrastructures for new scientific communities and builds upon earlier European and National projects, the purpose of the standardization task in MAPPER is to both instigate and coordinate the standardization efforts among the different project partners in support of applications that require and deploy resources and services in multiple infrastructures.

For the scope of this document we define a “standard” to be a commonly agreed, repeatable way of doing something. When using standards or adhering to standards we thus refer to published documents that contain a technical specification (or other precise criteria) designed to be used consistently as a rule, a guideline, or a definition.

Standards help to increase the reliability and the effectiveness of MAPPER services. They are designed for voluntary use and do not impose any regulations. However, laws and regulations may refer to certain standards and make compliance with them compulsory. Any standard is a collective work by manufacturers, users, research organizations, governmental departments and providers to achieve consensus on rules and their applicability. As standards are in principle “conceptware”, they need to be instantiated for specific situations. A major implication of this approach is the need for determining any implementation’s compliance with the agreed standards.

The standards considered in MAPPER (and thus in this document) relate to creating and using multiscale applications over European e-infrastructures. They fall into the following categories:

- information schemas defining the structure and meaning of information to be passed between peer services

- protocol specifications defining the messages that pass between peers and the way how peers interact
- Application Programming Interfaces (API) defining programming language interfaces that expose desired service functionalities within a client program

2.2 Methodology to Derive the Standardization Roadmap

The starting point for deriving the MAPPER Standardization Roadmap is represented by the MAPPER software stack as depicted in Figure 2.

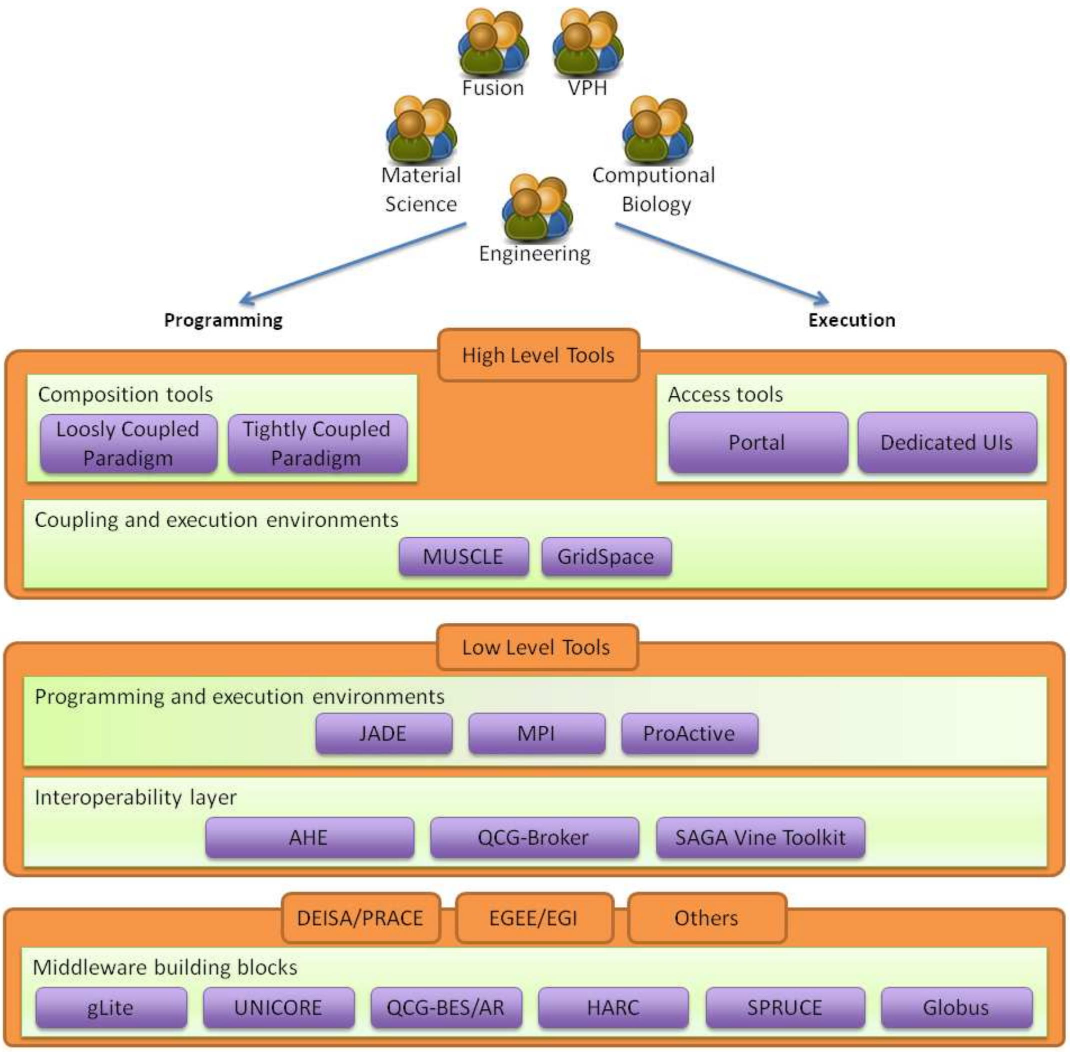

Figure 2: MAPPER software stack as described in MAPPER Deliverable D4.1 (Review of Applications, Users, Software and e-Infrastructures) [1].

From a long-term standardization/sustainability perspective, however, Figure 2 represents just one *specific* instantiation of the more abstract stack outlined in Figure 3, serving as a generic reference model for subsequent discussions.

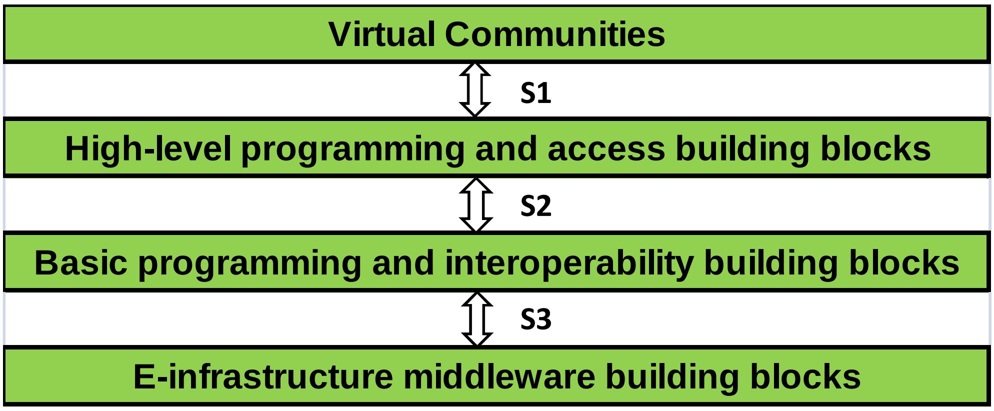

Figure 3: Abstract MAPPER stack.

The model is based on a layered approach with (heterogeneous) e-infrastructures on the bottom, a basic programming and interoperability layer in the middle, and a high level programming and access layer on top. Access to these layers is via the interfaces S1 to S3.

In more technical terms, a corollary from Section 2.1 is the necessity to standardize the interfaces S1 to S3 by answering the following questions

1. Which information schemas are necessary?
2. Which protocol interfaces (messages) are necessary?
3. Which APIs are necessary?

In order to answer these questions and translate them into a Roadmap, a thorough understanding of current standardization efforts as far as they relate to MAPPER is necessary.

2.3 Standardization Bodies with Relevance to MAPPER

As MAPPER has been setup as an interdisciplinary endeavour addressing both the application communities (fusion, hydrology, physiology, nano-material, computational biology) and the Grid ICT community, several standardization bodies contribute to MAPPER and vice versa. MAPPER aims at participating at these bodies during the runtime of the project and beyond. The intended impacts of MAPPER's contributions to the standardization bodies are a participation in the bodies' review processes and the evangelization on MAPPER standardization requests.

E-infrastructure related standards and standardization bodies have been well established during the last decade. They either refer directly to Grids or they impact the state-of-the-art in Grid activities. Appendix B provides more details of the e-infrastructure standards and standardization bodies relevant to MAPPER.

2.4 Methodology to Derive the Sustainability Plan

The MAPPER Sustainability Plan is described here in its first version. It will be revised over the course of the project. The objective of this Plan is to promote the sustainability of the technology and applications developed during the MAPPER project beyond the end of the project. Essentially, sustainability is concerned with the questions of ensuring that the solutions developed in the project will firstly continue to be available and functional for existing and new users; secondly support further developments; and thirdly provide ongoing training and troubleshooting activities.

The sustainability of the MAPPER solutions will thus depend on the long-term availability of MAPPER components, services and infrastructures; on long-term standards; and on long-term accounting and billing mechanisms.

The methodology to derive the Sustainability Plan is based on an analysis of the e-IRG Roadmap 2010 [2] as it describes the fundamental contribution of research e-infrastructures to Europe's competitiveness. Figure 4 illustrates an example of the importance of e-infrastructures for ICT innovations (like those the MAPPER project is aiming at) based on experiences and observations from the Dutch GigaPort project[1] [2].

The MAPPER Sustainability Plan follows a similar approach as it considers a basic layer of resources (network resources, computational resources, storage resources), a software layer with low level basic services and high level compound services building on the resource layer, an ICT Research and Development layer comprising methods and architectures, and finally a science layer devoted to pure scientific activities as considered in MAPPER work packages 7 and 8. Figure 5 summarizes these layers.

---

[1] http://www.surfnet.nl/en/innovatie/gigaport/Pages/Default.aspx

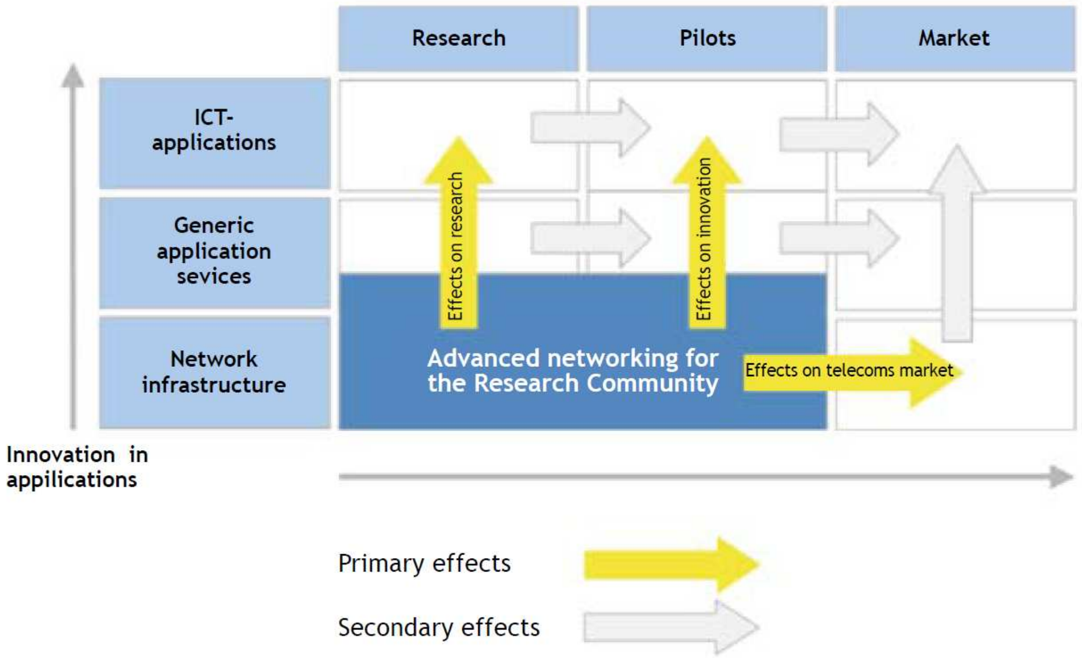

Figure 4: Example of the impact of infrastructure innovation [2].

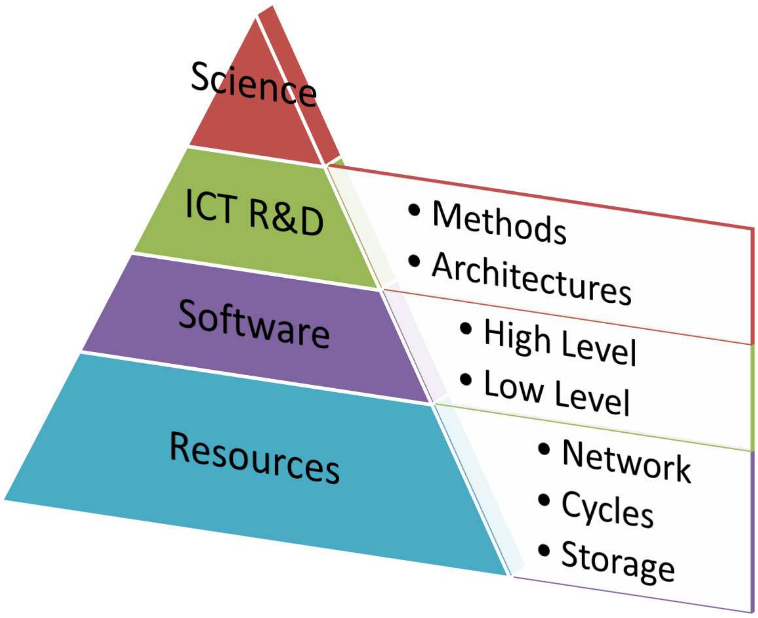

Figure 5: MAPPER sustainability pyramid.

for training, for support, for resource and service deployment, for tools, and for resource and service operation. As these latter categories can be mapped onto MAPPER work packages, the Sustainability Plan is canonically represented by the matrix in Figure 6. The subsequent discussions follow the layer dimension of Figure 5 and Figure 6.

		Operation / Deployment		Services	Dissemination, Training, Support			Policies, Standards, Interoperability	
		Independent Provider	VRC Sites	Tools Deployment	Dissemination	Training	Support	Policies	Standards & Interoperability
Science	WP7	VRC-driven, new curricula teaching science from an e-science perspective. MAPPER supports depoyment/usage of multi-scale modeling and simulation tools, including MAPPER tools.							
		WP6	WP4 / WP5	WP4 / WP5 /WP7	WP2	WP2	WP4 / WP8	WP3	WP3
ICT R&D	Methods Architecture	-	-	-	- School - Publications - WS/conference	School	-	-	-
Software	High level	-	Target specific VRCs: fusion, VPH, biology	- Sustained by partners	- D2.2: Diss. Plan - D2.4n: Diss. Reps - Publications - WS/conference	- School - Training materials	Yes* (at a cost)	- Support MAPPER-related VRC - EGI vs PRACE	- Use standards - Use OS-agnostic software
	Low level	EGI/PRACE Evaluation proc.	Plan B	-			Yes**.	- D3.1 → e-IRG, Comm. agencies - After project: ?	- Yes, support using OGF - Active involvement OASIS
Resources	Network Cycles Storage	-	Statistics in D6.2.	-	Statistics published in papers, on website, etc.	-	-	- Allocation policies - Resource allocation procedures - SLA, gSLA - e-IRG, GÉANT - Policy management board - EGI vs. PRACE policy	- Allocation standards - Collect best practises - D6.1 w.r.t. security

* Support is provided by MAPPER partners through e.g. project membership. We are considering to adopt other business models. In addition, we plan to integrate our software with the EGI infrastructure.
** The bulk of the software will be available as open source. The QCG layer is supported by PSNC. We will have to analyse this for the full software stack. (see also D5.1 and D6.1).

Key:

WP1	Management	VRC:	Virtual research communities	e-IRG:	e-Infrastructure Reflection Group
WP2	**Dissemination and Outreach**	Dess. Plan:	Dissemination Plan	OGF:	Open Grid Forum
WP3	**Policy Support and Sustainabil**	School:	MAPPER School	OASIS:	Open Standards for Info Society
WP4	**Adaptation of Existing Services**	WS:	Workshop	EGI:	Open Grid Infrastructure
WP5	**Vertical Integration**	OS:	Operating system	PRACE:	Advanced Computing in Europe
WP6	**E-Infrastructure Operations**				
WP7	**Application Enabling and Evaluation**				
WP8	**Programming and Execution Tools**				

Figure 6: Diagrammatic summary of the MAPPER Sustainability Plan.

3 Resources

Following Foster et al. [3], (computing) *resources* in the broader sense of this deliverable are physical resources like network elements, computing elements and storage elements. Such computing resources are “owned” by *resource providers* (i.e., organizations which are legal entities) and provided to VOs for temporary use subject to policies and constraints. Consequently, the perceived (virtual) life cycle of resources is coupled to the life cycle of VOs and VO membership of resource users – once the VO is decommissioned, the resources are no more available to the (previous) members, unless they join another VO which grants new access rights. While VOs represent structures implemented in the e-infrastructure for the access to resources that span multiple administrative and geographical boundaries, VRCs are collaborations of “like-minded” individuals that work in the same discipline, share a common interest or computational model which may span multiple VOs [4]. Typically, VOs are created and managed in VRCs.

MAPPER multiscale applications inherently rely on various resources of various types from various providers for various VOs in various VRCs [1, 4]. Examples include HPC systems provided for the Partnership for Advanced Computing in Europe (PRACE) and located in different countries, the data communications infrastructure provided by DANTE, and the storage capabilities available through EGI and the national grid initiatives.

Because MAPPER – as a project – will not provide any resources (in the sense of this report) to VOs, MAPPER is completely decoupled from pure resource provisioning by respective resource providers. On the other hand, the intention of MAPPER is to sustainably provide multiscale services. Consequently, arrangements have to be made on the resource level for enabling long-term access to required resources, both for members of VOs (whose life span outlives the MAPPER project) and for members of VOs whose establishment is scheduled after the project’s end.

Resource sustainability needs to be considered from different angles: from an operational perspective, a service perspective, an educational perspective, and an interoperability perspective according to Figure 6.

3.1 Operational/Deployment View

Assessing the sustainability of resources in the operational/deployment view requires differentiating between resources provided by “independent” resource providers (i.e., those who neither contribute to the MAPPER project directly nor indirectly by providing resources to the VRCs associated to MAPPER) and resources provided by VRCs associated with MAPPER.

As discussed previously, MAPPER will neither provide any resources to VOs by itself, nor will MAPPER be allowed to operationally manage resources outside the ownership of MAPPER partners (unless explicitly allowed by the resource owners). Thus, sustaining any resources (of independent resource providers and VRCs) is beyond MAPPER's scope and operational possibilities. Despite this general restriction, however, there is a difference between independent resources and VRC resources. While the life cycles of the former are completely independent from any MAPPER activities, the life cycles of the latter may have a relationship with MAPPER which is typically externalized as either

1. a direct project contribution (e.g., parts of the cluster at the Leibniz Supercomputing Centre), or
2. a service-level agreement (SLA) formally specifying the terms and conditions of guaranteed resource provisioning, or
3. a memorandum of understanding (MoU) describing an agreement with the VRC sites (acting as resource providers) for a convergence of will with an intended common line of action related to resource provisioning.

Note that these externalizations are time dependent: While a direct contribution (point 1) is only possible at project runtime, all other externalizations (points 2 and 3) may also be relevant after the project's end. The final version of this report will list the available resources in detail.

From a sustainability (of resources) point of view MAPPER will therefore instigate a series of MoUs and SLAs with Resource and Infrastructure Providers for long-term and dependable resource provisioning. A first MoU has been signed with EGI[2]; MAPPER has initiated also an evaluation process with PRACE. After the first evaluation step MAPPER middleware components have been accepted for testing procedures on selected PRACE sites, in particular SARA in Amsterdam, Netherlands.

---

[2] http://www.mapper-project.eu/documents/10155/25657/EGI-InSPIRE-MOU-MAPPER-FINAL.pdf

Long-term cooperation agreements with American initiatives like Xsede[3] and ExTENCI[4] are just as desirable as with Asian and Australian ones. Depending on the project results, MAPPER may provide as a separate deliverable a sample SLA to be used for after-project resource provisioning. Note, however, that such SLAs need to be associated with cost models, accounting and billing.

In order to simplify the agreements of SLAs and MoUs (at project runtime and post-project), MAPPER will keep an inventory of all deployed services in the public deliverable D6.2. This inventory may also serve as a checklist when arranging multi-SLAs with different providers. Furthermore, the operational statistics collected in D6.2 provide a good basis for the quantity structures SLAs and MoUs will typically rely on.

3.2 Service View

Because this section discusses sustainability of (physical) resources a separate service view is not necessary. Instead we refer to Section 4 for a more detailed analysis.

3.3 Educational View

The educational view refers to three separate aspects:

- dissemination and exploitation of results
- training in using results
- support while using results

3.3.1 Dissemination and Exploitation of Results

A sustainable provisioning of resources for MAPPER applications not only requires the resources themselves (see Section 3.1). Rather it requires the *right* resources. In order to attract resource providers to contribute their resources, they need to know the basic resource requirements. As pointed out previously, these will be collected in the public MAPPER Deliverable D6.2. In addition, MAPPER will publish the statistics on the MAPPER website (at least the public part of them) and in the context of scientific publications in dedicated journals [5].

[3] https://www.xsede.org/

[4] https://sites.google.com/site/extenci/

3.3.2 Training in Using Results

MAPPER will not provide training on using physical resources. Thus, this section is not applicable.

3.3.3 Support While Using Results

MAPPER will not support physical resources. Thus, this section is not applicable.

3.4 Interoperability View

The interoperability view not only covers standards and adjacent interoperability issues but also policies to adhere to.

3.4.1 Policies

Regarding the sustainability of resources, the policy aspect of the interoperability view needs to address two questions:

1. How can resources be allocated?
2. Which kinds of SLAs are required?

3.4.1.1 Resource Allocation

Resource allocation policies differ from provider to provider. A generic allocation description is therefore impossible. Rather we refer to the specific allocation procedures as outlined by the individual providers. As of this writing the following procedures are applicable and mandatory when sustainably allocating resources. Note, however, that most of these procedures are subject to changes without further notice to MAPPER. Thus, applicants are strongly advised to check the respective latest version. Further versions of this document will investigate changes as well.

Resource allocation in GÉANT

- GÉANT is the multi-gigabit pan-European research network providing connections between the networks of its European Research and Education Networks (NREN) partners. However, GÉANT does not connect individual researchers directly – rather, it does so indirectly through a "network of networks". Data sent by one end facility to another one would typically be sent across several local and regional networks before it reaches GÉANT, which would provide the international connection between the sending end facility and the receiving one. At the receiving end data flow again through several local and regional networks in the reverse order to that described before.

- NREN networks are connected to GÉANT via access links the speed of which depends on the respective NREN subscription, the capacity requirements of the NREN, and the widely differing availability and prices of connectivity across Europe.
- DANTE, the coordinator of the GÉANT project (not the network), also offers a so called DANTE World Service (DWS) for connectivity to the wider Internet.
- For more information we refer to the GÉANT home page[5].
- GÉANT provides inter-country connectivity for Europe's NRENs only, not for end users. From resource sustainability point of view, allocating a network resource requires to adhere to the national NREN procedures as for instance exemplified in the procedures of the German DFN-Verein[6].

Resource allocation in PRACE

- In order to assist parties interested in using PRACE resources, the PRACE consortium has prepared a comprehensive Application Guide. It describes in detail the various aspects of preparing a proposal requesting PRACE resources in response to calls for proposals, how the proposal is handled after submission, including peer reviews and resource allocation.
- The Application Guide is available from the PRACE home page[7].
- Applicants should also be aware that the efficient use of PRACE systems requires a detailed knowledge of architecture-specific factors influencing the performance (e.g., compilers, tools, libraries). PRACE therefore offers a set of best-practice-booklets on how to achieve good performance on these

---

[5] http://archive.geant.net/server/show/nav.159

[6] http://www.dfn.de/en/services/dfninternet/

[7] http://www.prace-ri.eu/Application-Guide-and-PRACE-Peer?lang=en

systems. These booklets are available on line[8] and it is strongly recommended to respect them.

Resource allocation in EGI

- EGI[9] denotes a pan-European Grid Infrastructure being established in close collaboration with the various European National Grid Initiatives (NGIs) and the European International Research Organisations (EIRO), to guarantee the long-term availability of a generic e-infrastructure for all European research communities and their international collaborators. As a matter of fact, EGI (or more precise EGI.eu as EGI's maintainer) does not own any resources; they still belong to and will be managed by the NGIs. EGI resource allocation is therefore formally externalized by a contract with and adhering to policies of the respective NGIs.
- However, since scientific research is expanding worldwide (the e-science paradigm), a longer lasting and easier operational model is needed − both for coordinating the e-infrastructure itself and for delivering integrated services that cross national borders. This is the motivation behind the EGI-InSPIRE project[10] which intends to support the transition from purely project-based systems to a sustainable pan-European e-infrastructure. EGI-InSPIRE will collect user requirements, provide coordination and support for current and emerging user communities, for example the ESFRI projects[11] (especially those with heavy usage demands like high energy physics, computational chemistry, or materials science – application areas fundamental to MAPPER), and define policies and procedures for pan-European resource allocation.
- From MAPPER's resource sustainability point of view these procedures need to be adopted and promoted as soon as they will be available.

[8] http://www.prace-ri.eu/Best-Practice-Guides?lang=en

[9] http://www.egi.eu/about/

[10] http://www.egi.eu/projects/egi-inspire/

[11] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

3.4.1.2 Service Level Agreements (SLA)

Once resource allocation procedures are finalized, MAPPER will as a deliverable of work package 6 provide sample versions of dedicated Services Level Agreements (SLA) for various constellations (see also Section 3.1) taking into account similar efforts in EGI.eu, PRACE, DANTE and national bodies (NREN, NGI). The MAPPER SLA samples will also leverage the work performed in the European gSLM project (Service Delivery and Service Level Management in Grid Infrastructures).[12]

3.4.2 Standards and Interoperability

Sustainably allocating, using and managing resources requires dedicated standards and proven compliances with these standards. While most of these standards (or de facto standards) will be defined by the parties mentioned in Section 3.4.1, others are and will be defined by more generically oriented standardization bodies like OGF, OASIS and similar ones.

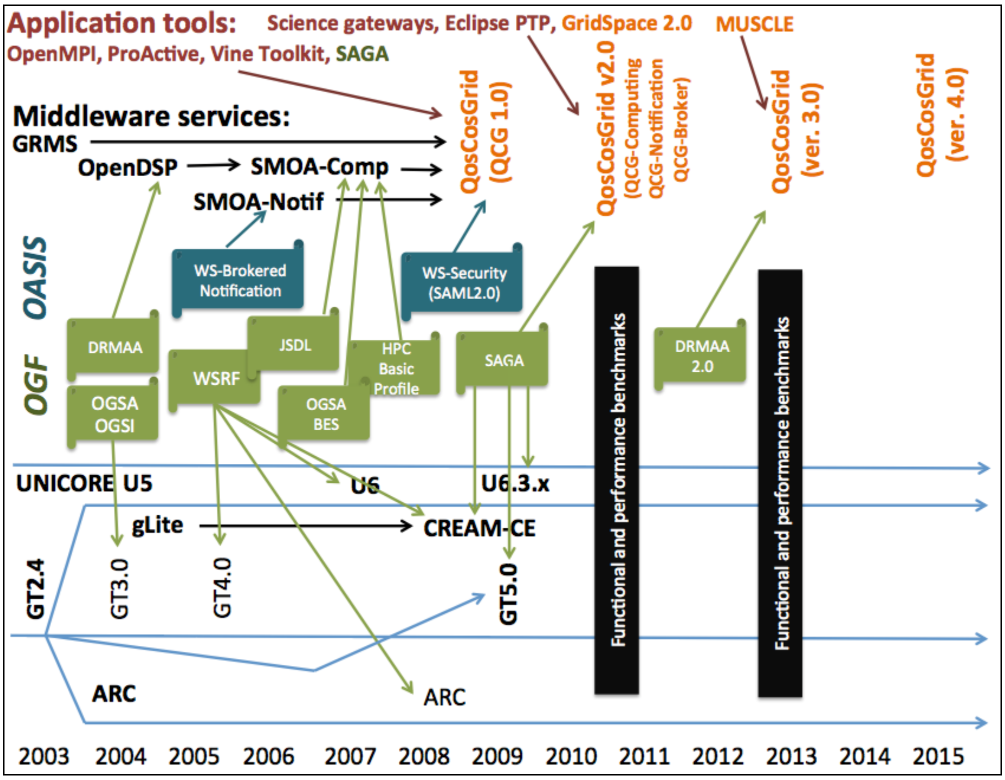

Figure 7: Middleware software evolution (QosCosGrid, UNICORE, gLite, ARC) and relevant OGF and OASIS standards

[12] http://gslm.eu/

A list of the most important OGF and OASIS interoperability standards that have been supported in MAPPER together with a roadmap are presented in Figure 7. In order to support the sustainable access to resources, MAPPER not only contributes to several standardization bodies (see Section 2.2 and Appendix A). MAPPER goes even a step beyond and defines a specific MAPPER profile with a corresponding test suite to sustainably determine the compliance of future applications to the profile. The profile will be the focus of deliverable D3.3; the test suite will be described in deliverable D3.5. Also, be aware that deliverable D6.1 (Report on the Assessment of Operational Procedures and Definition of the MAPPER Operational Model [6]) contains material to take into account when reasoning about sustainable access to resources.

4 Software

The process of creating the MAPPER software and integrating it vertically with existing e-infrastructures has been based on the predefined reusable software components since the beginning, see Figure 1. The MAPPER consortium has contributed with many legacy software components that need to be integrated and maintained onto e-infrastructures to support new or better capabilities offered to users and their distributed multiscale applications. From the sustainability perspective it is important to understand that MAPPER distinguishes two software creation phases, fast-track and deep-track respectively. During the fast-track phase we have already reused available software components and integrated them into a consistent middleware based on the QosCosGrid v2.0, see Figure 7. We believe that twofold advantages will be gained thanks to such a software development and sustainability strategy:

- a systematic development of reusable components, and
- a systematic reuse of MAPPER components as building blocks to create new application scenarios.

In fact, by a reusable component we not only mean the code. We see even bigger benefits of reusing MAPPER software specifications, reference implementations of standards, application use cases or templates. The fast-track components deliver core functionalities needed, especially in case of tightly coupled multiscale applications that must be executed and managed on co-allocated computational resources across many sites in the e-infrastructure. However, whenever possible, the

MAPPER project will use components already deployed on the e-infrastructure and recommended by EGI and PRACE. From sustainability and maintenance points of view we can group MAPPER related components into four groups:

1. local domain (low level) MAPPER services (QCG-Computing – an advance reservation and execution service, QCG-Notification - a notification service)
2. grid domain (high level) MAPPER services and user interfaces (QCG-Broker – a load balancing and co-allocation service, Application Hosting Environment)
3. multiscale coupling libraries (MUSCLE for tightly coupled multiscale applications, GridSpace for loosely coupled multiscale applications)
4. third party e-infrastructure services maintained and supported by external communities (UNICORE, gLite or Globus Toolkit)

The first group of software components belong to the fast-track phase, whereas the other two groups belong to the deep-track phase. Third party e-infrastructure services are out of scope of the MAPPER sustainability plan as external communities support them directly. Nevertheless, as it was mentioned in previous sections, MAPPER has already delivered the software to EGI and PRACE for the evaluation processes under established joint task forces.

The deep-track software creation and integration phase will begin during the M12-M36 period of the MAPPER project and at this point is not directly included within the software sustainability plan. One should note that high-level services and application tools that are part of the deep-track phase are not directly integrated with e-infrastructure services and less efforts is required within Service Activities to support the software as it will be deployed in the user-space. In other words, once the access to new capabilities in MAPPER is provided and supported to end users, they can modify and maintain their own software depending on their requirements.

To conclude, we believe that our sustainability approach adopted in the fast-track phase being fully based on proven software reuse could:

- increase software productivity,
- shorten software development time during the next deep-track phase,
- improve software system interoperability,
- reduce software development and maintenance costs,

- produce more standardized and better quality software.

Similarly to resources, sustainability of software needs to be considered from different angles: from an operational perspective, a service perspective, an educational perspective, and an interoperability perspective according to Figure 6. In the next subsections we will try to address those aspects of sustainability by describing various strategies reflecting different class of software components.

4.1 Operation and Deployment View

This section briefly describes the MAPPER software from both operation and deployment perspectives. The common feature of middleware services is that they must be deployed and then supported locally by a system administrator of a certain e-infrastructure site, whereas application tools can be deployed by users that have the authorized access to the site (see Table 1).

Table 1: Software capabilities

Software	Capabilities						
Middleware services:	Basic Jobs	MPI Jobs	Loosely coupled Jobs	Tightly coupled Jobs	Advance Reservation support	Automated notifications about Jobs	Automated co-allocation support
QCG- Computing	X	X	X	X	X		
QCG-Notification						X	
QCG-Broker	X	X	X	X	X	X	X
AHE	X	X	X		X		
UNICORE	X	X	X				
gLite	X		X				
Application tools:							
GridSpace	X	X	X		X	X	X
MUSCLE		X		X	X		X
SAGA / Vine Toolkit	X	X	X		X		

It is worth mentioning that some MAPPER software components may require additional privileges (like the job submission and advance reservation service), other just one inbound ports open, etc. Therefore, some additional deployment and administrative efforts are required on the site. However, all the detailed configuration descriptions, templates and examples have been also provided according to requirements defined by EGI and PRACE.

As the MAPPER project aims to deploy a computational science environment for distributed multiscale computing, on and across European e-infrastructures, it has initiated already a collaboration with PRACE (PaRtnership for Advanced Computing in Europe) and with EGI-InSPIRE (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe). During the first collaboration meeting between the MAPPER, EGI and PRACE project teams held in May 2011, it was determined that as a first step, two applications should be integrated over the forthcoming period to perform distributed multiscale computing on the selected sites of production e-infrastructure. In our opinion this approach has two advantages. First, it can be used to promote MAPPER within EGI and PRACE communities. Second, the MAPPER software components will be evaluated by external entities. The carefully selected applications address major needs of two main multiscale application use cases in MAPPER, namely tightly and loosely coupled scenarios. The coordination of activities relevant to this collaboration will be performed via an expressly formed MAPPER-PRACE-EGI Task Force[13] comprising specialists from each of the three organisations.

Almost all e-infrastructures shape their own policies and procedures regarding installing new software. Those procedures usually require operational and security audits of every component to be deployed on the production e-infastructure. It is worth mentioning that most of the components already passed successfully such procedures in the national grid initiatives. Currently a subset of MAPPER components is being evaluated within PRACE.

---

[13] MAPPER-PRACE-EGI Task Force wiki: https://wiki.egi.eu/wiki/MAPPER-PRACE-EGI_Task_Force_(MTF)

4.2 Services View

Within the Service Activity work packages we have already published the D6.1 Report on the Assessment of Operational Procedures and Definition of the MAPPER Operational Model document that presents the services view on the MAPPER software [6]. It discusses the assessment of operational procedures of the European e-Infrastructures targeted by MAPPER, including DEISA, EGI and PRACE, and defining the operational model for MAPPER software.

4.3 Dissemination, Training and Support View

The software employed in the MAPPER project benefits from the fact that various software components provided by the consortium have been matured and supported over the last few years, see Figure 7. Consequently, we were able to setup core middleware services, including QCG-Computing, QCG-Notification and QCG-Broker, onto selected production EGI sites in NGI PL-Grid within the first six months of the project as it was reported in the MAPPER D5.1 Report on the Inventory of Deployed Services. This successful effort has a direct impact on other NGIs in EGI and resulted in the internal task force within EGI to support MAPPER. Thus, some of dissemination, training and support procedures will be adopted by MAPPER in the near future.

Additionally, many training and dissemination activities for different software components have been already performed under the EU funded projects:

- QosCosGrid – developed since 2005, supported in QosCosGrid, BREIN and PL-Grid projects, disseminated during various conferences talks, interoperability events and publications. Currently the QosCosGrid stack is deployed on the majority of PL-Grid production sites.
- GridSpace – developed since 2006, supported in Virolab, Gredia, PL-Grid projects, disseminated in tutorials held at the conferences, several conference and journal publications, with ongoing training activities addressed to Polish scientific communities.
- AHE - developed since 2005, supported in Virolab, OMII-Europe and RealityGrid projects.

It is worth mentioning that the contracted support within the PL-Grid project is a 5 year-long period that reaches 2017. Till this time both GridSpace and QosCosGrid will be available for end users (including support). As indicated, the original compounds of the Mapper project have been and will be disseminated and supported (including user support). On the other hand the compilation of already existing original elements enhanced with specific multiscale applications' demands need to be addressed as a whole and presented to multiscale applications' communities as a comprehensive environment dedicated to them.

After the first prototype deployment of newly designed multiscale application tools (month 12) is ready, the new activities have to be carried out to spread knowledge about it, and preferably, engage alpha, beta tester and allow early-access for interested users. This early dissemination would help in collecting of users from the very beginning and help to collect immediate feedback which is not to underestimate while validating the system and planning the roadmap for the solution. In the context of the second prototype (month 24) it seems even more vital to relate to the feedback collected by that time. The sound and solid foundation of the consortium partner institutions and their obligations to support the software and their users in next several years as well as the already sustainable and disseminated efforts related to this software make a good path to follow in the scope of the MAPPER project.

4.4 Policies View

From the MAPPER software perspective, especially for many-cluster multiscale model executions, the most crucial are site networking policies that concerns network firewall configuration. Those network policies usually differ between sites; however a minimal set of requirement is the following:

- all traffic is allowed between worker nodes and interactive node,
- interactive node allows outgoing connection to the external sites,
- upon special agreement some inbound ports can be opened on the interactive node.

For this reason all communication and networking channels established between single scale modules spread out on different e-Infrastructure Grid/HPC sites must be routed via transport overlay deployed on the interactive node. The detailed description of networking policies will be reported in the D5.2 MAPPER vertical integration plan.

4.5 Standards and Interoperability View

Many Grid/HPC middleware technologies have been augmented with implementations of proposed standards from the Open Grid Forum, see Figure 7. From the middleware interoperability perspective, the key specification is the OGSA Basic Execution Service (BES) for job management and submission. By supporting BES it is possible to improve the interoperability between different Grids or HPC environments as well as allow upper level services and application tools, e.g. SAGA, Vine Toolkit or GridSpace, to access efficiently remote computational resources in a uniform way. In fact, BES provides easy, intuitive and standardized access to computational resources. The BES specification defines two mandatory (BESFactory and BESManagement) and one optional (BESActivity) interface. The management interface allows controlling the service itself. The factory interface provides job submission and bulk monitoring capabilities while the activity interface allows monitoring of a single job. All the jobs submitted to the BES interface have to be described in the Job Submission and Description Language (JSDL), another relevant OGF standard from the interoperability perspective. Finally, we believe that there are also many efforts required at the high-level application programming to bridge the gap between existing grid middleware and application-level needs. The Simple API for Grid Applications (SAGA) is an OGF standardization effort that addresses this particular gap by providing a simple, stable, and uniform programming interface that integrates the most common grid programming abstractions. These most common abstractions were identified through the analysis of several existing and emerging applications, including multiscale application scenarios. As it is presented in Figure 6, all relevant grid middleware services commonly used in the e-Infrastructure, including gLite, UNICORE, Globus Toolkit and more importantly QosCosGrid, are support key aforementioned standards for interoperability.

Therefore, as a proof of concept, we have already performed interoperability and performance tests in MAPPER to access remotely job submission and monitoring capabilities provided by the recent stable and recommended grid middleware services for e-Infrastructures, namely gLite 3.2, UNICORE 6.4, Globus Toolkit 5.0.3 and QosCosGrid 2.0. The first set of benchmarking tests measured the throughput of

1000 jobs, described according to the JSDL standard, and then submitted via high-level SAGA libraries to underlying grid middleware services via BES interfaces. As it is presented in Figure 8, the QosCosGrid middleware outperforms other available grid middleware services by about order of magnitude. A similar throughput benchmarking tests but for 50 jobs concurrently submitted by 10 users only proved our assumptions that the QosCosGrid middleware adopted in MAPPER is the most efficient implementation of interoperability standards. All our benchmarks have been performed on selected EGI production sites available for MAPPER users. Detailed descriptions will be included in appropriate Service Activity deliverables and technical reports managed by OGF.

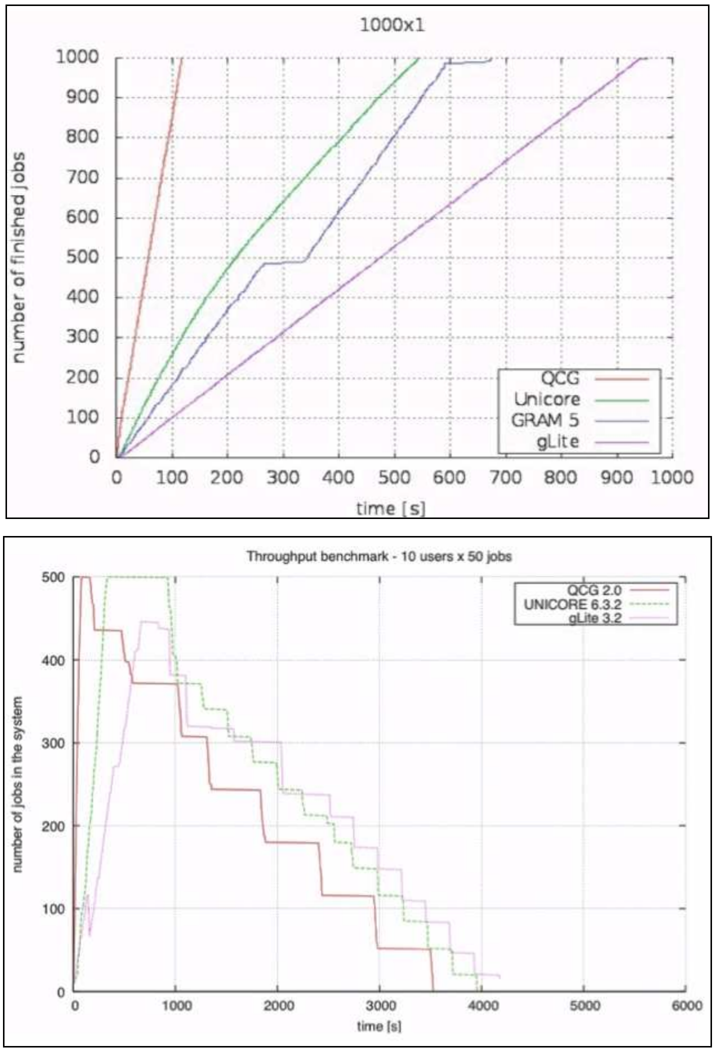

Figure 8: The overall performance of the QosCosGrid BES implementation compared to UNICORE, gLite, and Globus.

We envision that further interoperability improvements will include tests of new capabilities in the e-Infrastructure, e.g., advance reservation defined by a new version of the OGF DRMAA standard released in 2011. Currently, the QosCosGrid is the only grid middleware supporting advance reservation and co-allocation. Moreover, we believe that there are still many efforts needed to improve the interoperability between different authentication and authorization processes.

Additionally, the MAPPER project, by the means of the WP3 activities, will use and promote standards on many different layers. The planned actions will be focused on:

- taking active participation in standardization bodies, in particular OGF and OASIS,
- adopting new standards in MAPPER components, such as OGF DRMAA, OASIS SAML2.0, etc.,
- providing feedback and own requirements on evolving standards,
- lobbing on deploying software components that adheres to open standards (a new task forces within EGI and PRACE have been already established to deal with this task),
- taking participation in interoperability events, like the OGF BES interoperability demo[14].

Finally, as the general rule, all software components delivered by MAPPER will be OS-agnostic software and will follow the commonly accepted software packaging guidance[15].

5 ICT R&D

Sustainability of ICT R&D within MAPPER concerns the scientific methods and architectures which are developed within the project. Both aspects are of a conceptual nature and therefore are best sustained by effectively spreading methods and architectures as applied in MAPPER to a wider community. This will be

[14] Interoperability Event at OGF30: http://forge.ogf.org/sf/go/wiki2458

[15] Fedora Packaging:Guidelines: http://fedoraproject.org/wiki/Packaging:Guidelines

accomplished by providing opportunities for training and by preparing high-quality publications for peer-reviewed journals and conferences.

- **Training**: MAPPER will organize at least two MAPPER Seasonal Schools. The first of these is scheduled for year two. Its aim is to train members of the project in installing and using MAPPER tools. The first Seasonal School will mainly be visited by consortium members plus a limited number of external participants. The second MAPPER Seasonal School, scheduled for year three, will be fully open to interested researchers, developers, scientists and other interested parties. In this edition of the Seasonal School all aspects of the project, including the MAPPER concepts, methods and tools, will be covered. Once the MAPPER tools have been deployed on the service stack of EGI (and later PRACE) the school may also be advertised through EGI and PRACE channels.
- **Publishing and presenting**: MAPPER members aim at regularly attending a wide range of scientific conferences and at publishing scientific methods and architectures in international journals and conference/workshop proceedings. The MAPPER Dissemination Plan [5] provides a detailed description of the publication approach.
- **Community building**: Because distributed multiscale computing is an emerging field, MAPPER focuses their sustainability efforts not only on effective dissemination and training but also on building a community for distributed multiscale computing. By contributing to a coherent and connected multiscale modelling and simulation R&D community, MAPPER aims at increasing the public awareness of (distributed) multiscale computing.

MAPPER's initial effort into bolstering an interdisciplinary multiscale community is based on organising an international workshop for Distributed Multiscale Computing at the IEEE e-Science conference.[16] The first edition of this workshop is due to take place in December 2011, a second workshop is planned towards the end of 2012. Note that these workshops will also be offered after the project's end.

---

[16] Workshop website: http://www.computationalscience.nl/dmc2011/.

6 Science

Part of MAPPER's approach to sustainability is to ensure that MAPPER concepts and solutions are used by a wide range of science and engineering domains for multiscale modelling and simulation. Since MAPPER's financial resources are limited, MAPPER is focusing on a small set of application areas including biomedical, nano-material science, hydrology and flow control and nuclear fusion. The principal approach consists of facilitating support and training for deploying and using DMC tools, including MAPPER tools, for scientific investigations in these areas. Among other things, this may be achieved by sourcing new funding for research projects in these areas. An additional element of sustainability is concerned with the development of new curricula teaching e-science, in particular with emphasis on multiscale modelling and simulation.

In the following we briefly outline aspects of sustainability in relation to the main scientific areas covered by the project.

- **HemeLB.** The work performed on HemeLB [8] within MAPPER is a starting point for a sustained evolution of HemeLB into a large multiscale method to resolve blood flow in human arteries. The emphasis on the code is to provide medically valid results, practical clinical assistance, and real-time visualisation. HemeLB has already been in use at UCL since 2007 and currently being enhanced.

  MAPPER introduces a hierarchical coupling of scales to efficiently model blood flow from the main artery scale all the way down to the capillaries. To ensure the sustainability of this tool, MAPPER is pursuing several directions. First of all, MAPPER aims at validating HemeLB for use in clinical environments through extensive testing in collaboration with the UCL Hospital (UCLH). Second, MAPPER will incorporate the resulting improvements directly into the HemeLB codebase. As a result, MAPPER will be able to reuse the results in several large projects which feature HemeLB as an example application. First, HemeLB will be part of the 3-year FP7 project CRESTA (Collaborative Research into Exascale Systemware, Tools and Applications), which starts in October 2011. Second, the code will be central in the EPSRC-funded project LSLBBC (Large Scale Lattice-Boltzmann simulation of BioColloids), which begins in January 2012. This particular project will fund a research at UCL and will involve enhancement of HemeLB and subsequent use. And third, the code will be adopted within a 3-year 2020 Science project which started in July 2011.

- **Material Sciences.** The rationale for sustaining multiscale approaches in MAPPER's nanomaterial applications focuses on adopting multiscale approaches to enable the modelling and simulation of more complex systems and to obtain a more robust statistical sampling of the properties of clay-polymer interactions.

  The main sustainability driver in the nanomaterial domain is the adoption of widely used and open-source tools. All simulation codes used within the nanomaterial application (LAMMPS, CPMD) are publicly available and are coupled using core MAPPER components such as GridSpace, AHE and QCG Broker. In addition to ensuring that all components are open-source, MAPPER aims at including user manuals for deploying and operating multiscale simulations using these codes. Although MAPPER will provide manuals and publish its methods, MAPPER will not be providing support on an individual user basis to communities.

  Additionally, MAPPER intends to apply its multiscale solutions in a project from the Qatar National Research Fund, named "From Fundamental Understanding to Predictive Design of Layered Nanomaterials", which began in December 2010 and which will run for 3 years.

- **Modelling of in-stent restenosis and cranial aneurysms.** The multiscale nature of biomedical systems is very well recognized in the physiology communities, as well as the need for consistent approaches to the coupling of sub-models where many of them need HPC facilities to be executed (see VPH framework). Within the EU-funded projects MeDDiCA and THROMBUS, the UvA is concerned with the modelling of In-stent restenosis and stent-induced thrombosis in cranial aneurysms.

  - THROMBUS[17] is an EU-funded FP7 collaboration of biomedical and computational science research institutions, hospitals, and a stent

[17] http://www.thrombus-vph.eu/

company that started in February 2011 and will run for 3 years. THROMBUS is concerned with the development and validation of a multiscale model for the stent-induced thrombosis in cranial aneurysms. Besides a better understanding of the processes involved in thrombosis, the ultimate goal is to develop a multiscale computational model and simulation framework that can be used in clinical environments to support clinicians in choosing the right stent and the most effective deployment strategy. Large-scale distributed multiscale computation facilities enable both, the thorough validation of a simplified multiscale model that can run locally in clinical environments, and the execution of the large-scale multiscale model for patient-specific data in an urgent computing scheme. The methodologies developed within MAPPER will be applied. MAPPER activities regarding policies and computing services are important steps to enable urgent biomedical multiscale computing in the future. The THROMBUS project will have direct impact on more predictive, effective and safer healthcare.

- MeDDiCA[18] is a Marie Curie Initial Training Network (MC ITN) focused on Cardiovascular Engineering and Medical Devices that started in September 2009 and will run for 4 years. Within MeDDiCA at the UvA we will further develop a computational multiscale model for in-stent restenosis (ISR). The ISR model has been developed during the FP6 COAST project that lay the foundation for many aspects of multiscale modelling and execution now brought to a next stage within MAPPER. The ISR model is one of the most advanced use cases in MAPPER and MAPPER tools will be directly implemented and validated using this model. Main objective within MeDiCCA is to improve the existing ISR model and enrich its predictive capabilities regarding the abnormal growth of smooth muscle cell tissue after stenting arteriosclerotic arteries. A number of biomedical, clinical, and computational science research institutions are involved in this project, as well as a simulation software company, all working on the modelling and understanding of

[18] http://www.meddica.eu/

cardiovascular diseases and the design of cardiovascular devices. Within this framework multiscale methodologies and simulation services developed within MAPPER are actively promoted and shared with the other partners.

The methodologies, software frameworks, and services developed within MAPPER facilitate the development and execution of these models and will be further promoted through the projects' dissemination activities and subsequent research collaborations. Through our involvement in the VPH community MAPPER methodologies and tools will eventually also find its way into the VPH tool kit, an intended extensive collection of biological sub-models and software frameworks to couple them.

More in general, the European wide VPH community has organized themselves into the VPH Institute. UvA and UCL are members of the institute and through the institute dissemination of MAPPER tools and services will be undertaken. Moreover, MAPPER will establish and foster contacts between the VPH institute and other e-infrastructure projects such as PRACE and EGI-InSPIRE, thus fostering further dissemination of MAPPER results.

- **Flow control.** Currently, the MAPPER project offers a promising framework to develop large multiscale applications. UNIGE is involved in the simulation of irrigation canals and other water courses. Flow control is an important challenge for computational science. Many practical problems require us to act on some part of a fluid flow. This is the case of the management of irrigation canal where gates should be adjusted dynamically to fulfil the water demand, as a function of the meteorological situation or the need of the users. Obvious constraints are to avoid the overflow or draining of the canal, or a waste of water. Other flow control problems are related to the accumulation of sediments in lakes, near canal gates, etc. By adjusting flow parameters, the deposit can be brought to adequate places.

  In a completely different context, blood flow control is becoming of central importance in many biomedical application (see also in-stent restenosis above), and in particular in cerebral aneurysms. Devices, called flow diverters (or stents) are inserted in the parent artery to control the flow in the aneurysm cavity and produce its occlusion.

- **Computational biology.** The University of Ulster focuses on applications in the fast-emerging field of systems biology. While many approaches to systems biology involve single-scale models, a growing body in recent research aims at modelling life phenomena across several scales. Multiscale systems biology is concerned with experiments and hypotheses that involve different scales of biological organization from intracellular molecular interactions to cellular behaviour and the behaviour of cell populations, to an organ or even a whole organism. While many models in systems biology require solutions that deal biological processes that at multiple space and/or time scales or different levels of biological organization, there is also a need to allow us to quantitatively explore biological systems by means of simulations that incorporate heterogeneous DMC techniques, and to scale the computation of simulations. To facilitate a long-term sustainability of MAPPER technology, UU focuses on standard, widely used computational biology tools and technology, including the SBML, the Systems Biology Results Markup Language (SBRML) and Copasi [9]. By focusing on these standards (or de facto standards), we hope to establish some of the concepts, methods and tools developed by MAPPER in systems biology community. We are also working on e-science teaching material for systems biology in which we want to emphasize DMC using MAPPER results.

7 Action Plan

In order to implement both the sustainability plan and the standardization roadmap MAPPER proposes an organisational setup and several dedicated actions. The organisational setup defines the (structural and procedural) context of the actions to be performed. During the project’s runtime the context is implicitly defined by the project’s structure. After the end of the project it is proposed to create a MAPPER Virtual Organisation (VO) with the objective to further promote the MAPPER ideas and results (action 12 in Table 2).

The primary sustainability actions are listed in Table 2. It should be noticed that these actions represent generic activities and thus need to be instantiated. For example, action 9 will refer to several standardization bodies. It should also be noted that the beneficiary leading the action is implicitly determined by the supporting deliverables (column (e)).

Table 2: Actions to implement the sustainability plan

Id	Action	Action to be performed		Primary source of information (if delivered during the project)	Primary source of information (if delivered after the project)
		during the project	after the project		
(a)	(b)	(c)	(d)	(e)	(f)
1	Provide a description of all necessary resources	x		D4.1, D6.2, D7.3, D8.1	
2	Provide a description of all necessary and available services and software packages	x	x	D5.1, D6.2, D8.1	Resource and Service Providers (using a MAPPER provided template)
3	Provide a user manual of how to get access to and use the resources and services	x	x	D8.4	Resource and Service Providers (addenda to the MAPPER manual)
4	Provide a support manual	x		D6.3	
5	Provide an SLA template	x		D6.2	
6	Negotiate MoUs	x		ongoing	
7	Present MAPPER results	x	x	D2.2, D2.3	MAPPER final results presentation
8	Provide MAPPER training material	x		D2.5.1, D2.5.2	
9	Representation in standardization bodies	x	x	ongoing	personal participation
10	Definition and promotion of MAPPER profile	x		D3.3, D3.5	
11	Publish scientific results obtained from applying MAPPER tools and services	x	x	ongoing	
12	Create a MAPPER VO consisting of MAPPER partners and interested stakeholders		x		to be defined
13	Apply for follow-on projects		x		to be defined

8 Conclusion

This document specifies the initial versions of a Sustainability Plan and a corresponding Roadmap. Both are required for ensuring that the results of the MAPPER project are adopted and further developed. The sustainability approach is structured into four layers of tasks and activities with each layer being associated with various standardization and sustainability categories. This document also defines an initial version of the MAPPER Standardization Roadmap. The Roadmap describes the standards relevant to ongoing MAPPER implementations and mid-term to long-term issues to be solved for e-infrastructures enabling multiscale applications.

This document is a living document which will be updated in month 24 (deliverable D3.6) based on the findings during the course of the project. Consequently, this initial version defines just the starting point. It is by purpose incomplete in several sections. The gaps will be filled by later versions.

9 References

[1] Deliverable D4.1 of the MAPPER project: Review of Applications, Users, Software and e-Infrastructures, 2011.

[2] e-Infrastructure Reflection Group (e-IRG): e-IRG Roadmap 2010. www.e-irg.eu/images/stories/eirg_roadmap_2010_layout_final.pdf

[3] Foster, I.; Kesselman, C.; Tuecke, S.: The Anatomy of the Grid: Enabling Scalable Virtual Organizations; *International Journal of High Performance Computing Applications,* 2001*, 15*, 200-222.

[4] e-Infrastructure Reflection Group (e-IRG): e-IRG White Paper 2011. http://www.surveymonkey.com/s/SP3WP2011

[5] Deliverable D2.2 of the MAPPER project: Final Plan for the Use and Dissemination of Foreground, 2010. www.mapper-project.eu/documents/10155/23424/D2.2+-+Dissemination+Plan.pdf

[6] Deliverable D6.1 of the MAPPER project: Report on the Assessment of Operational Procedures and Definition of the MAPPER Operational Model, 2011. www.mapper-project.eu/documents/10155/23478/D6.1+-+Operational+Procedures+and+Model.pdf

[7] Cossu, R., Petitdidier, M., Linford, J., Badoux, V., Fusco, L., Gotab, B., Hluchy, L., Lecca, G., Murgia, F., Plevier, C., Renard, P., Schwichtenberg, H., Som de Cerff, W., Tran, V., Vetois, G.: A Roadmap for a Dedicated Earth Science Grid Platform. Earth Science Informatics 3 (3), 2010, pp. 135-148

[8] Mazzeo, M.D., Coveney, P.V. (2008) HemeLB: A high performance parallel lattice-Boltzmann code for large scale fluid flow in complex geometries. Comput Phys Commun, 178 (12) 894-914.

[9] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. & Kummer, U.: COPASI – a COmplex PAthway Simulator, Bioinformatics, 22(24):3067–3074, 2006.

10 Abbreviations

API	Application Programming Interface
BES	Basic Execution Service
BioPAX	Biological Pathway Exchange
DANTE	Delivery of Advanced Networking Technology to Europe
DICOM	Digital Imaging and Communcations in Medicine
DMC	Distributed Multiscale Computing
DMTF	Distributed Management Task Force
EGI	European Grid Infrastructure
EGI-InSPIRE	EGI Integrated Sustainable Pan-European Infrastructure for Researchers in Europe
e-IRG	e-Infrastructure Reflection Group
EIRO	European International Research Organisation
ESFRI	European Strategy Forum on Research Infrastructures
ExTENCI	Extending Science Through Enhanced National Cyberinfrastructure
FieldML	Field Modelling Markup Language
FTP	File Transfer Protocol
GO	Gene Ontology
gSLM	Service Delivery and Service Level Management in Grid Infrastructures
ICT	Information and Communication Technology
IETF	Internet Engineering Task Force
IRI	Internationalized Resource Identifier
MAPPER	Multiscale Applications on European e-Infrastructures
MITA	Medical Imaging and Technology Alliance
MoU	Memorandum of Understanding
MMS	Multiscale modelling and simulation
NEMA	Association of Electrical and Medical Imaging Equipment Manufacturers
NGI	National Grid Infrastructure
NREN	National Research and Education Network
OASIS	Organization for the Advancement of Structured Information Standards
OBO	Open Biological and Biomedical Ontologies

OGF	Open Grid Forum
OGSA	Open Grid Services Architecture
OMG	Object Management Group
PRACE	Partnership for Advanced Computing in Europe
SBGN	Systems Biology Graphical Notation
SBML	Systems Biology Markup Language
SIENA	Standards and Interoperability for e-Infrastructure Implementation Initiative
SLA	Service Level Agreement
VO	Virtual Organization
VRC	Virtual Research Community
W3C	World Wide Web Consortium
WP	Work Package
WSDL	Web Service Description Language
XML	Extended Mark-Up Language
Xsede	Extreme Science and Engineering Discovery Environment

Appendix A: Standardization Roadmap

Introductory Note

To instigate a wider impact MAPPER services and applications needs to be deployable over several existing e-Infrastructures (PRACE, EGI, etc.) and future (potentially heterogeneous) ones. MAPPER therefore adheres to and contributes to standards required for achieving interoperability and sustainability (see Section 2.3 of this document). The discussions in Section 2.3 imply an inventory of standards for the interfaces S1, S2 and S3 (see Figure 3), serving as starting point for the Roadmap towards a standards-based system for multiscale solutions below.

Roadmap

Criteria for a Standardized e-Infrastructure for Distributed Multiscale Computing

As discussed in other MAPPER deliverables ([1, 5, 6]), the main improvements to be achieved by standardizing e-infrastructures for multiscale applications are the following:

- simplification of access to infrastructure elements like networks, computing elements, storage elements by easy to use APIs and core middleware services for retrieving data, storing real time data, scheduling jobs in a heterogeneous and federated environment; for fast transfer of large files and a large number of different files;
- ease of management of metadata by user/role based access control and by support for special types of databases beyond the relational model and SQL queries (like e.g., spatial and column databases);
- reliability improvement of processing combined with efficient Quality of Service (QoS) and fault tolerance mechanisms, with advance reservation techniques, and with pre-emption of calculations for e.g. operational jobs or contingence reactions due to risk alerts;
- harmonized authentication and authorization mechanisms for usage in portals;
- easy integration with application-specific web portals;

- easy incorporation into business-driven scenarios by providing mechanisms for monitoring, accounting and billing.

General Objectives

The Roadmap described below aims at defining the key steps necessary for providing a standardized e-Infrastructure for multiscale applications.

The Roadmap is adapted from similar efforts (like [7] which itself is based on roadmaps for EGI, CoreGrid, e-IRG, and others). However, compared to these works, our focus is on multiscale applications in e-infrastructures (especially Grids) rather than on Grid deployments themselves. Although there are several similarities, both roadmaps differ as the one presented here is intended to be implemented (at least in parts) by standardization bodies. Hence, MAPPER contributes to the various bodies; however, MAPPER will not – and cannot – prescribe any standards themselves.

The macro objectives of the Roadmap are intended to achieve the final overall objective, i.e. a standardized multiscale application e-infrastructure. They are assumed to be realized in a stepwise manner, involving several -- often transversal -- goals. For example, all objectives require some kind of formal description of resources for identifying them or for increasing the number of available resources. Similarly, all objectives require solving the access issues for resource-aware multiscale applications or the ability to migrate multiscale applications to other infrastructures.

Following [7], the Roadmap is arranged along some key categories, specifying their expected achievements and the perceived time period (short-term, mid-term, long-term).

Short-Term Objectives

One of the major objectives to be addressed short-term is awareness building for the specific standardization requirements implied by MAPPER. To achieve this objective, several activities have been started or are being started. Examples are the work performed by the MAPPER work package WP2 (Dissemination) and the activities defined in EGI-InSPIRE.

As awareness building (in the above sense) has a strong governance component, for an e-infrastructure focusing on multiscale applications to achieve efficiency of scale, many organizations need to participate (by granting access to scientific equipment, to Grid nodes, to high performance computing systems, etc.). Without VRC specific directives this will probably not happen. For this reason we urge the key actors on the political and funding levels to engage in a lasting dialogue to address the key issues raised by this Roadmap proposal. An important instrument for this is deliverable D3.1 on policies, which has been and still is actively distributed with key actors, in order to raise awareness on the consequences of requirements of DMC on the e-Infrastructures, and to get such issues on the political agendas. Closely related to awareness building activities are specialized e-Infrastructure training programs focusing specifically on multiscale issues.

Apart from building awareness, multiscale-aware science communities require technical and operational benefits from standardization. This includes (but is not limited to) single sign-on authentication to all resources and services, policy-based authorization of accessing restricted data (support for user roles, directory services, etc.), transparent migration of Grid jobs upon availability of “better” resources or as part of global fail-over strategies, transparent adaptation of middleware services, or support for heterogeneous programming paradigms.

For a successful deployment of multiscale applications at large, the identification and discovery of tools and services across communities is needed. A short-term goal is thus a standardized way not only to describe, to address and to discover multiscale services but also to describe, to address and to discover competence centres specializing on multiscale applications.

It should be noted that in parallel to this top-down approach, a demand-driven bottom-up approach is necessary as well. Demonstrations of any kind (videos, live demonstrations, booths at fairs, etc.) can create interest and momentum for standardization of “multiscaleness” which in turn may result in an elevated interest in other standardization areas. This intertwined demand generation for standardization needs to be backed up short- and medium-term by a number of initiatives and the development of respective components in all building block areas, by the creation of fascinating demonstrators, and by claiming compliance to a standards based MAPPER profile as proposed in deliverable D3.3 and deliverable D3.5 for the corresponding compliance test suite.

Finally, it is worth noting that at the time of writing this report, some actions related to these short-time objectives have already been taken and are ongoing. Several special sessions related to the standardization of multiscale-conform e-infrastructures were held at conferences and workshops (e.g. OGF). An important outcome of these activities was the general request to ensure the standards are interoperable between diverse standardization bodies.

The following table summarizes some of the short-term objectives.

Short-term objectives	• to be achieved within the next 3 years
Expected Achievements	• expand the VRCs for multiscale applications in terms of number of organisations, number of projects, number of applications • reduce the technological gaps by providing a set of formal service and resource descriptions, a set of interface descriptions and a set of basic and higher coupling services • port multiscale-aware applications to comply with standards (initially for disseminations purposes, but gradually moving to production-type applications)
Means (examples)	• creation of synergies between various VRCs and continuous promotion and dissemination activities (e.g., at specific conferences) • definition of a first version of the APIs for various building blocks • definition of a first draft of metadata tools • dissemination and exploitation

Mid-Term Objectives

The main mid-term concerns are solving the outstanding blocking issues for multiscale applications to be deployed on an e-Infrastructure, the porting of existing multiscale applications from other disciplines, and enabling the “citizen scientist” to benefit from multiscale applications. The main mid-term objectives are summarized in the table below.

Mid-term objectives	• to be achieved within the next 2 to 5 years
Expected Achievements	• Identification of and solutions to blocking issues for multiscale applications (e.g., availability of resources) • including data access and • quality of service • Supporting new multiscale applications (especially from other research disciplines) • to use the standards • publication of scientific results which could not have been achieved without standardization
Means (examples)	• increase the number of multiscale applications to use the standards and let them benefit immediately for having access to sensors, data, archives, etc. • expose and publish the interfaces and standards • define a MAPPER profile multiscale applications should comply with (both upon development and upon deployment) • definition of an enhanced version of APIs • define dynamic authentication and authorization policies for accessing data and resources to assure Quality-of-Service levels • define methods for easy deployment of multiscale applications • dissemination and exploitation

Long-Term Objectives

The long-term objectives are fundamental for the sustainability aspects of MAPPER. However, due to the time range involved, they can only be recommendations which cannot be achieved as long as critical mid-term objectives are still open. The main long-term objectives are summarized in the table below.

The effort required for achieving these objectives is more about end user engagement than it is about the deployment of any particular technology. Nonetheless, it should be kept in mind that the development of standard-compliant buildings blocks and their deployment is not an easy task which will require some effort. Consequently, special emphasis should be given long-term to the ease of adoption, the ease of use, the hiding of complexity from the user, and the support of more complex scientific workflows relying multiscaleness.

The following table summarizes the long-term objectives.

Long-term objectives	• to be achieved within the next 4 to 10 years
Expected Achievements	• Establish a standardized e-infrastructure to deploy multiscale applications into which fully integrates the defined standards and which complies with the profile being defined mid-term
Means (examples)	• definition of the dedicated e-Infrastructure • make multiscale application services and service data and models readily available via infrastructure services • establish within 10 years a framework to provide timely data of interoperable applications for local, national, regional, and international policy makers

Appendix B: Standardization Bodies and Standards

E-Infrastructure related standardization bodies

E-Infrastructure related standardization bodies have been well established during the last decade. They either refer directly to grids or they impact the state-of-the-art in grid activities.

Open Grid Forum (OGF)

> "OGF[19] (Open Grid Forum) is an open community committed to driving the rapid evolution and adoption of applied distributed computing like Grid 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."[20]

The OGF is organized in several research and community groups. The ones most relevant for MAPPER are

- Grid Interoperation Now Community Group (gin-cg)
- Remote Instrumentation Services in Grid Environment - RG (risge-rg)
- Storage Networking Community Group (sn-cg)
- Grid Information Retrieval RG (gir-rg)
- Firewall Issues RG (fi-rg)
- Grid High-Performance Networking RG (ghpn-rg)
- Grid Scheduling Architecture RG (gsa-rg)
- Infrastructure Services On-Demand Provisioning Research Group (ISOD-RG)
- Levels of Authentication Assurance Research Group (loa-rg)

In addition there are standards functions responsible for the development of architectures, specifications, roadmaps and glossaries for distributed computing software through OGF working groups. The most relevant ones for MAPPER are:

---

[19] http://www.ogf.org

[20] Quote from http://www.gridforum.org/About/abt_overview.php

- Access to Remote Instrumentation in a distributed environment – Working Group (ari-wg)
- Certificate Authority Operations WG (caops-wg)
- Data Format Description Language WG (dfdl-wg)
- Database Access and Integration Services WG (dais-wg)
- Distributed Computing Infrastructure Federation Working Group (dcifed-wg)
- Distributed Resource Management Application API WG (drmaa-wg)
- Firewall Virtualization for Grid Applications WG (fvga-wg)
- GLUE Working Group (glue)
- Grid File System Working Group (gfs-wg)
- Grid Remote Procedure Call WG (gridrpc-wg)
- Grid Resource Allocation Agreement Protocol WG (graap-wg)
- Grid Storage Management WG (gsm-wg)
- GridFTP WG (gridftp-wg)
- High Performance Computing Profile WG (hpcp-wg)
- Job Submission Description Language WG (jsdl-wg)
- Network Mark-up Language Working Group (nml-wg)
- Network Measurement and Control WG (nmc-wg)
- Network Measurements Working Group (nm-wg)
- Network Service Interface WG (nsi-wg)
- OGSA Authorization WG (ogsa-authz-wg)
- OGSA Basic Execution Services WG (ogsa-bes-wg)
- OGSA ByteIO Working Group (byteio-wg)
- OGSA Data Movement Interface WG (ogsa-dmi-wg)
- OGSA Naming Working Group (ogsa-naming-wg)
- OGSA Resource Selection Services WG (ogsa-rss-wg)
- Open Cloud Computing Interface WG (occi-wg)
- Open Grid Services Architecture WG (ogsa-wg)
- Production Grid Infrastructure WG (pgi-wg)
- Simple API for Grid Applications WG (saga-wg)

MAPPER will also establish a close relationship to the European branch of OGF by participating at the SIENA initiative (Standards and Interoperability for e-Infrastructure Implementation Initiative)[21].

Organization for the Advancement of Structured Information Standards (OASIS)

OASIS[22] (founded in 1993) is a not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society. The consortium produces several Web services standards necessary for the usage of Grid services as envisioned by MAPPER. The following OASIS standards are relevant for MAPPER:

- eXtensible Access Control Markup Language (XACML)
- Reference Model for Service Oriented Architecture
- Security Assertion Markup Language (SAML)
- Service Provisioning Markup Language (SPML)
- Web Services Notification (WSN)
- Web Services Resource Framework (WSRF)
- Web Services Security
- Web Services Transaction
- Web Services ReliableMessaging
- WS-Trust

Distributed Management Task Force (DMTF)

DMTF[23] enables more effective management of IT systems worldwide by bringing the IT industry together to collaborate on the development, validation and promotion of systems management standards. With its deep and broad reach, DMTF creates standards that enable interoperable IT management as necessary for MAPPER work package 6. DMTF management standards are critical to enabling management interoperability among multi-vendor systems, tools and solutions within the enterprise. The DMTF standards stack may be derived from Figure 9.

---

[21] http://www.sienainitiative.eu/StaticPage/About.aspx

[22] http://www.oasis-open.org

[23] http://www.dmtf.org

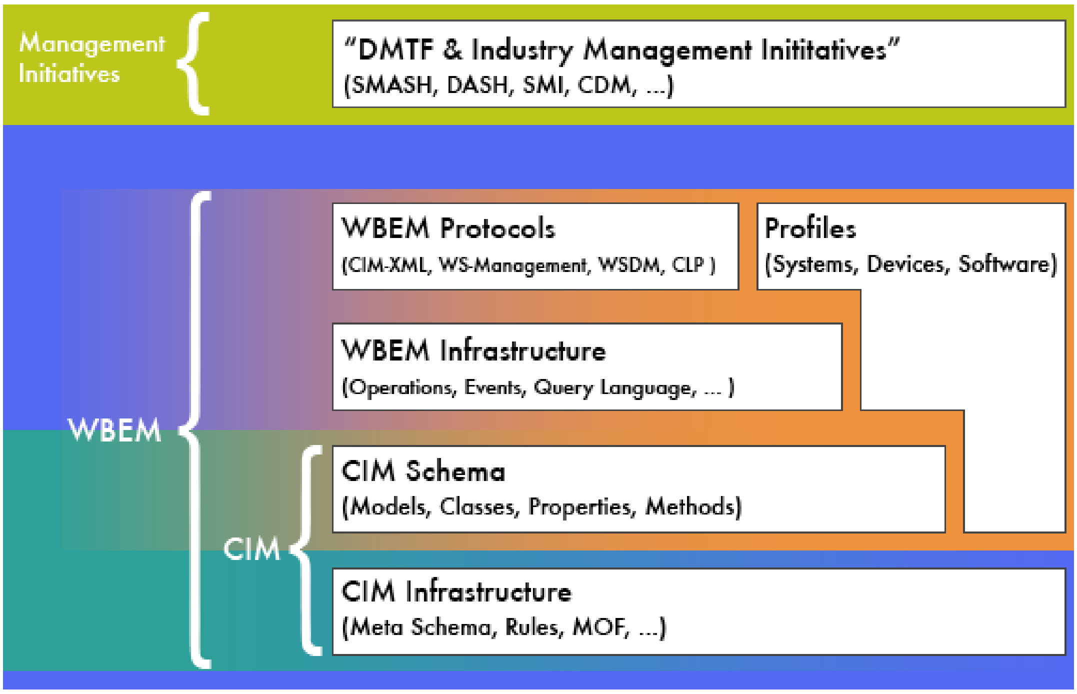

Figure 9: DMTF Standards overview.[24]

As MAPPER applications will technically rely on commonly agreed information models, a close relationship with the respective DMTF standardization work would be beneficial.

World Wide Web Consortium (W3C)

The W3C[25] primarily pursues its mission through the creation of Web standards and guidelines. Since 1994, W3C has published more than 110 such standards, called W3C Recommendations. In order for the Web to reach its full potential, the most fundamental Web technologies must be compatible with one another and allow any hardware and software used to access the Web to work together. W3C refers to this goal as *Web interoperability*. By publishing open (non-proprietary) standards for Web languages and protocols, W3C seeks to avoid market fragmentation and thus Web fragmentation. W3C "owns" the MAPPER-relevant standards for XML, WSDL, and SOAP.

---

[24] http://xml.coverpages.org/dmtf-cim.html

[25] http://www.w3.org

Internet Engineering Task Force (IETF)

The mission of IETF[26] is to produce high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet. These documents include protocol standards, best current practices, and informational documents of various kinds. Due to its distributed nature MAPPER has to be built on IETF standards like constrained RESTful environments, FTP extensions, Internationalized Resource Identifiers (IRI), or Web Security.

Object Management Group (OMG)

OMG[27] has been an international, open membership, not-for-profit computer industry consortium since 1989. OMG Task Forces develop enterprise integration standards for a wide range of technologies like real-time, embedded and specialized systems, analysis and design, architecture-driven middleware and others. The relevance for MAPPER is imposed by OMG's model driven application development approach.[28]

Most of the standards prepared and maintained by these standardization bodies relate to each other (e.g., OGSA standards are based on OASIS ones which rely on W3C ones). They represent best practices in Grid environments and up to certain degrees they have been implemented compliantly in various Grid middleware solutions (Globus Toolkit, UNICORE, gLite). As a consequence, this report will not address each standard separately. Rather we will elaborate a more general pattern for MAPPER.

Application-related standardization bodies and standards

MAPPER applications fall into the categories fusion, hydrology, physiology, nano-material and computational biology [1]. In order to develop and execute applications in these categories, MAPPER looks at standards, quasi standards and resources discussed in the bodies and organizations like (see Table 3 for references to relevant websites):

---

[26] http://www.ietf.org

[27] http://www.omg.org

[28] http://www.omg.org/mda/

- SBML.org maintains the Systems Biology Markup Language (SBML) standard.
- BioPAX.org maintains the Biological Pathways Exchange (BioPax) standard.
- SBGN.org maintains the Systems Biology Graphical Notation (SBGN) standard.
- There are also several ontology bases to consider as a quasi standard: (a) the Gene Ontology (GO) Consortium provides ontologies of defined terms representing gene product properties; and (b) the Open Biological and Biomedical Ontologies (OBO) support the development and publication of ontologies in the biomedical domain.
- The BioModels Database is a repository of peer-reviewed, published, computational models.
- The Medical Imaging and Technology Alliance (MITA), a division of the Association of Electrical and Medical Imaging Equipment Manufacturers (NEMA), maintains the standard on Digital Imaging and Communications in Medicine (DICOM).
- The goal of the FieldML (Field Modelling Markup Language) is to support the building of hierarchical models represented by generalized mathematical fields in a standardized way.
- The CellML language is an open standard maintained by the Auckland Bioengineering Institute at the University of Auckland and affiliated research groups. The purpose of CellML is to store and exchange computer-based mathematical models.

Table 3: Web addresses of relevant standards and resources.

Standard, Resource	Website URL
SBML	http://sbml.org/Documents/Specifications
CellML	http://www.cellml.org/specifications
BioPAX	http://www.biopax.org/specification.php
FieldML	http://www.physiome.org.nz/xml_languages/fieldml
SBGN	http://www.sbgn.org/Documents/Specifications
GO	http://www.geneontology.org/

BioModels	http://www.ebi.ac.uk/biomodels-main/
MITA	http://www.medicalimaging.org
NEMA	http://www.nema.org
DICOM	http://medical.nema.org/