EUROPEAN MIDDLEWARE INITIATIVE

DNA1.4 - EMI ROADMAP AND DCI COLLABORATIONS

EC DELIVERABLE: D1.4

Abstract:
This deliverable describes the planned EMI roadmap and the collaboration objectives within the European and international Distributed Computing Infrastructures. In particular it places EMI with the general context of the European Infrastructure vision and describes specific common objectives and collaboration mechanism with other relevant infrastructure, middleware and support projects.
### Delivery Slip

<table>
<thead>
<tr>
<th>From</th>
<th>Partner / Activity</th>
<th>Date</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alberto Di Meglio</td>
<td>CERN/NA1</td>
<td>07/12/2010</td>
<td></td>
</tr>
</tbody>
</table>

| Reviewed by      |                     |            |           |
| Emidio Giorgio   | INFN/NA2 LU/JRA1     | 16/12/2010 |           |
| Oxana Smirnova   |                     |            |           |

| Approved by      |                     |            |           |

### Document Log

<table>
<thead>
<tr>
<th>Issue</th>
<th>Date</th>
<th>Comment</th>
<th>Author / Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>14/07/2010</td>
<td>Table of Content</td>
<td>Alberto Di Meglio</td>
</tr>
<tr>
<td>0.2</td>
<td>14/11/2010</td>
<td>Added content to roadmap sections, added DCI Collaboration Roadmap document</td>
<td>Alberto Di Meglio</td>
</tr>
<tr>
<td>0.3</td>
<td>24/11/2010</td>
<td>Added collaborations with external projects and initiatives (DCI, OSG, WLCG, etc)</td>
<td>Alberto Di Meglio</td>
</tr>
<tr>
<td>0.3</td>
<td>07/12/2010</td>
<td>Released for internal review</td>
<td>Alberto Di Meglio</td>
</tr>
<tr>
<td>0.31</td>
<td>08/12/2010</td>
<td>Fixed headers and footers</td>
<td>Alberto Di Meglio</td>
</tr>
<tr>
<td>0.4</td>
<td>27/02/2011</td>
<td>Revision after reviewers’ comments</td>
<td>Alberto Di Meglio</td>
</tr>
<tr>
<td>1.0</td>
<td>05/03/2011</td>
<td>Final version</td>
<td>Alberto Di Meglio</td>
</tr>
</tbody>
</table>

### Document Change Record

<table>
<thead>
<tr>
<th>Issue</th>
<th>Item</th>
<th>Reason for Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>Clarified goals and format of the deliverables in the Introduction and Executive Summary</td>
<td>Address reviewers’ comments</td>
</tr>
<tr>
<td></td>
<td>Added synoptic table of high-level milestones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Added more information about the major goals of the EMI major releases</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>Final version for release</td>
<td></td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

1 INTRODUCTION ..................................................................................................................6
  1.1 PURPOSE .........................................................................................................................6
  1.2 DOCUMENT ORGANISATION .........................................................................................6
  1.3 REFERENCES ..................................................................................................................7
  1.4 DOCUMENT AMENDMENT PROCEDURE ....................................................................7
  1.5 TERMINOLOGY ...............................................................................................................8

PART 1

1 EXECUTIVE SUMMARY ......................................................................................................12
2 THE EMI ROADMAP ..........................................................................................................13
  2.1 INTRODUCTION .............................................................................................................13
  2.2 THE EMI STRATEGIC VISION .......................................................................................14
  2.3 THE EMI ROADMAP AT A GLANCE ............................................................................14
  2.4 THE EMI SOFTWARE DEVELOPMENT ROADMAP .......................................................17
    2.4.1 The yearly development cycle ...............................................................................20
  2.5 THE EMI COLLABORATION ROADMAP .......................................................................22
3 THE DCI PROJECTS AND OTHER EUROPEAN COLLABORATIONS ...............................24
  3.1 EGI-INSPIRE ................................................................................................................24
  3.2 OTHER DCI PROJECTS ................................................................................................25
  3.3 PRACE AND OTHER HIGH-PERFORMANCE COMPUTING INITIATIVES ..........................25
  3.4 WLCG ..........................................................................................................................26
  3.5 EEF ..................................................................................................................................26
  3.6 ESFRI PROJECTS .........................................................................................................26
  3.7 STANDARDIZATION BODIES .........................................................................................27
4 OSG AND OTHER INTERNATIONAL COLLABORATIONS ...............................................28
  4.1 OPEN SCIENCE GRID ..................................................................................................28
  4.2 LATIN AMERICA COLLABORATIONS .........................................................................28
  4.3 ASIA-PACIFIC COLLABORATIONS ..............................................................................29
5 INDUSTRIAL COLLABORATIONS .......................................................................................30
6 CONCLUSIONS ....................................................................................................................31

PART 2

1 EXECUTIVE SUMMARY ......................................................................................................33
2 INTRODUCTION ..................................................................................................................34
3 A VISION FOR EUROPEAN DCIS ...................................................................................35
  3.1 WHAT ARE DCIS? .........................................................................................................35
  3.2 A DECADE OF COMMUNITY BUILDING .....................................................................35
  3.3 CURRENT CHALLENGES ..............................................................................................36
  3.4 EGI AND THE DCI EVOLUTION ....................................................................................37
  3.5 THE DCI COMMON VISION ..........................................................................................38
  3.6 IMPLEMENTING THE VISION........................................................................................39
    3.6.1 For Infrastructure Providers ..................................................................................39
    3.6.2 For the Software Provider .....................................................................................40
4.6 VENUS-C ................................................................. 48
  4.6.1 EGI-InSPIRE ..................................................... 49
  4.6.2 EMI ........................................................................ 49
  4.6.3 IGE ....................................................................... 49
  4.6.4 EDGI ..................................................................... 49
  4.6.5 StratusLab ........................................................... 50

5 CONCLUSIONS .............................................................. 51

6 ANNEX A: THE DISTRIBUTED COMPUTING INFRASTRUCTURE PROJECTS ...... 52
  6.1 EUROPEAN GRID INFRASTRUCTURE – INTEGRATED SUSTAINED PAN-EUROPEAN INFRASTRUCTURE FOR RESEARCHERS IN EUROPE (EGI-INSPIRE) .............................................................. 52
  6.2 EUROPEAN MIDDLEWARE INITIATIVE (EMI) ................................................................. 53
  6.3 INITIATIVE FOR GLOBUS IN EUROPE (IGE) ................................................................. 54
    6.3.1 Objectives .............................................................. 54
    6.3.2 Action plan ............................................................ 55
  6.4 EUROPEAN DESKTOP GRID INITIATIVE (EDGI) .......................................................... 56
  6.5 STRATUSLAB ............................................................ 57
    6.5.1 Summary .............................................................. 57
    6.5.2 Objectives .............................................................. 57
    6.5.3 Action plan ............................................................ 57
  6.6 VIRTUAL MULTIDISCIPLINARY ENVIRONMENTS USING CLOUD INFRASTRUCTURES (VENUS-C) .... 58
1 INTRODUCTION

1.1 PURPOSE

This document describes the overall EMI Roadmap and its planned collaborations with the European DCI Projects and other European and international infrastructures and research projects. This document is a high-level roadmap document of the EMI strategy providing the general context for the other more detailed EMI planning deliverables. It consists of two main parts: the first part is a description of the EMI Development Roadmap, its collaboration ecosystem and a set of strategic positioning statements with respect to other projects and initiatives in Europe and world-wide. The second part is the overall European DCI Projects Roadmap elaborated jointly with EGI-InSPIRE, IGE, EDGI, VENUS-C and StratusLab. This second part is common to all DCI Projects and is included in this deliverable without modifications as agreed among the DCI Projects representatives. It puts EMI into the context of the global DCI vision, clarifying the role of EMI in achieving the vision and the specific opportunities for collaboration with each of the DCI Projects.

1.2 DOCUMENT ORGANISATION

This document is organized as follows:

Chapter 1: Introduction: this section, explaining the purpose, scope and organization of the document

Part 1: The EMI Development Roadmap and its Strategic Positioning in the Context of European and International Collaborations

Chapter 1: Executive Summary: This section contains a high-level description of Part 1 document with some references to Part 2. It gives a summary of the most important points described in each main section.

Chapter 2: The EMI Roadmap: This section describes the overall EMI strategic development and collaboration roadmap and milestones over its three year duration, the expected inputs and outputs and the role played by the collaboration with other project and initiatives in the roadmap implementation.

Chapter 3: The DCI Projects and other European Collaborations: This section describes the common collaboration program established among the six so-called DCI Projects and other European infrastructure projects. It provides updates and additional information, not available during the preparation of the common roadmap document.

Chapter 4: OSG and Other International Collaborations: This section describes the collaboration with the Open Science Grid initiative in the US, initiatives in the Asia-Pacific area and possible plans for further extensions.

Chapter 5: Industrial Collaborations: This section describes the collaboration with industrial partners.

Chapter 6: Conclusions: This section provides information on expected further work to implement the EMI roadmap.

PART 2: Distributed Computing Infrastructure (DCI) Collaborative Roadmap

Chapter 1: Executive Summary: This section contains a high-level description of Part 2 document. It gives a summary of the most important points described in each main section.
Chapter 2: Introduction: this section introduces the six DCI projects and describes the context for their collaboration.

Chapter 3: A Vision for European DCIs: this section describes the vision elaborated by the six DCI projects together and the most important expected outcome of the next several years.

Chapter 4: Individual Interactions: this section describes in a comprehensive way the interactions and the opportunities for collaboration among the individual projects to implement the agreed vision.

Chapter 5: Conclusions: Conclusions and expected impact of the DCI collaboration.

Annex A: The Distributed Computing Infrastructure Projects: a description of each DCI project objectives and program of work.

### 1.3 REFERENCES

| R1 | DNA2.1.1 - EMI Collaboration Programs v. 1.0 | http://cdsweb.cern.ch/record/1277553?ln=en |
| R2 | DNA2.3.1 - Dissemination and Use of Knowledge Plan v. 1.0 | http://cdsweb.cern.ch/record/1277621?ln=en |
| R3 | DNA2.2.1 - Training Plan v. 1.0 | http://cdsweb.cern.ch/record/1277575?ln=en |
| R4 | DNA2.4.2 - Exploitation and Sustainability Plan | http://cdsweb.cern.ch/record/1277605?ln=en |
| R5 | DSA1.1 - Software Maintenance and Support Plan v. 1.0 | http://cdsweb.cern.ch/record/1277556?ln=en |
| R6 | EMI Project Description of Work | https://twiki.cern.ch/twiki/pub/EMI/EmiDocuments/EMI-Part_B_20100624-PUBLIC.pdf |
| R7 | DNA1.3.1 – Technical Development Plan | http://cdsweb.cern.ch/record/1277540?ln=en |
| R8 | DSA1.2 - Software Release Plan v. 1.0 | http://cdsweb.cern.ch/record/1277545?ln=en |
| R10 | DNA1.2.1 - EMI Service Level Agreement Template v. 1.0 | http://cdsweb.cern.ch/record/1277517?ln=en |
| R11 | DJRA1.5.1 - Standardization Workplan and Status Report v. 1.0 | http://cdsweb.cern.ch/record/1277526?ln=en |

### 1.4 DOCUMENT AMENDMENT PROCEDURE

This document can be amended by the EMI Project Director or Technical Director or people appointed by them to this task further to any feedback from other teams or people. Minor changes, such as spelling corrections, content formatting or minor text re-organisation not affecting the content and meaning of the document can be applied by the EMI Project Director or Technical Director without
peer review. Other changes must be submitted to peer review and to the EMI PEB and TCB for approval.

When the document is modified for any reason, its version number shall be incremented accordingly. The document version number shall follow the standard EMI conventions for document versioning. The document shall be maintained in the CERN CDS repository and be made accessible through the OpenAIRE portal.

### 1.5 TERMINOLOGY

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARC</td>
<td>The Advanced Resource Connector is a general purpose, Open Source, lightweight, portable middleware solution (<a href="http://www.knowarc.eu/middleware.html">http://www.knowarc.eu/middleware.html</a>)</td>
</tr>
<tr>
<td>CHAIN</td>
<td>Project co-funded by the European Union in the 7th Framework Program to gather experience from Regional eInfrastructures in several continents.</td>
</tr>
<tr>
<td>dCache</td>
<td>System for storing and retrieving huge amounts of data, distributed among a large number of heterogeneous server nodes, under a single virtual filesystem tree with a variety of standard access methods (<a href="http://www.dcache.org">http://www.dcache.org</a>)</td>
</tr>
<tr>
<td>DCI</td>
<td>Distributed Computing Infrastructure</td>
</tr>
<tr>
<td>DEISA</td>
<td>The Distributed European Infrastructure for Supercomputing Applications, is a consortium of leading national Supercomputing centres that aims at fostering the pan-European world-leading computational science research (<a href="http://www.deisa.eu">http://www.deisa.eu</a>)</td>
</tr>
<tr>
<td>EDGI</td>
<td>European Desktop Grid Infrastructure (<a href="http://edgi-project.eu">http://edgi-project.eu</a>)</td>
</tr>
<tr>
<td>EEF</td>
<td>European EInfrastructures Forum (<a href="http://www.einfrastructures-forum.eu">http://www.einfrastructures-forum.eu</a>)</td>
</tr>
<tr>
<td>EELA</td>
<td>E-science grid facility for Europe and Latin America (<a href="http://www.eu-eela.eu">http://www.eu-eela.eu</a>)</td>
</tr>
<tr>
<td>EGI</td>
<td>European Grid Infrastructure – <a href="http://www.egi.eu">http://www.egi.eu</a></td>
</tr>
<tr>
<td>ESFRI</td>
<td>The European Strategy Forum on Research Infrastructures, is a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach (<a href="http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri">http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri</a>)</td>
</tr>
<tr>
<td>EU-IndiaGrid</td>
<td>EU-IndiaGrid support specific user communities in the exploitation of grid infrastructure in areas strategic for EU-Indian collaboration (<a href="http://www.euindiagrid.eu">http://www.euindiagrid.eu</a>)</td>
</tr>
<tr>
<td>GEANT</td>
<td>GÉANT2 is the high-bandwidth, academic Internet serving Europe’s research and education community (<a href="http://www.geant2.net">http://www.geant2.net</a>)</td>
</tr>
<tr>
<td>GGUS</td>
<td>Global Grid User Support (<a href="https://gus.fzk.de">https://gus.fzk.de</a>)</td>
</tr>
<tr>
<td>GISELA</td>
<td>Grid Initiatives for e-Science virtual communities in Europe and Latin America aims at implementing the Latin American Grid Initiative (LGI) sustainability model rooted on National Grid Initiatives (NGI) or Equivalent Domestic Grid Structures (EDGS), in association with CLARA, Latin American NRENs and collaborating with the European Grid Initiative (EGI) (<a href="http://www.gisela-grid.eu">http://www.gisela-grid.eu</a>)</td>
</tr>
<tr>
<td><strong>gLite</strong></td>
<td>The next generation middleware for grid computing born from the collaborative efforts of more than 80 people in 12 different academic and industrial research centers as part of the EGEE Project (<a href="http://glite.web.cern.ch/glite/">http://glite.web.cern.ch/glite/</a>)</td>
</tr>
<tr>
<td><strong>IGE</strong></td>
<td>Initiative for Globus in Europe (<a href="http://www.ige-project.eu">http://www.ige-project.eu</a>)</td>
</tr>
<tr>
<td><strong>KnowARC</strong></td>
<td>&quot;Grid-enabled Know-how Sharing Technology Based on ARC Services and Open Standards&quot; (KnowARC) is a Sixth Framework Programme Specific Targeted Research Project, under Priority IST-2005-2.5.4 &quot;Advanced Grid Technologies, Systems and Services&quot;. The project began in June 2006 and ends in November 2009 (<a href="http://www.knowarc.eu">http://www.knowarc.eu</a>)</td>
</tr>
<tr>
<td><strong>NorduGrid</strong></td>
<td>A Grid Research and Development collaboration aiming at development, maintenance and support of the free Grid middleware, known as the Advance Resource Connector (ARC) (<a href="http://www.nordugrid.org/">http://www.nordugrid.org/</a>)</td>
</tr>
<tr>
<td><strong>NSF</strong></td>
<td>The National Science Foundation (NSF) is a United States government agency that supports fundamental research and education in all the non-medical fields of science and engineering (<a href="http://www.nsf.gov/">http://www.nsf.gov/</a>)</td>
</tr>
<tr>
<td><strong>OASIS</strong></td>
<td>A not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society (<a href="http://www.oasis-open.org/home/index.php">http://www.oasis-open.org/home/index.php</a>)</td>
</tr>
<tr>
<td><strong>OGF</strong></td>
<td>An open community committed to driving the rapid evolution and adoption of applied distributed computing (<a href="http://www.gridforum.org/">http://www.gridforum.org/</a>)</td>
</tr>
<tr>
<td><strong>OSG</strong></td>
<td>The Open Science Grid brings together computing and storage resources from campuses and research communities into a common, shared grid infrastructure over research networks via a common set of middleware (<a href="http://www.opensciencegrid.org/">http://www.opensciencegrid.org/</a>)</td>
</tr>
<tr>
<td><strong>PRACE</strong></td>
<td>Partnership for Advanced Computing in Europe, a unique persistent pan-European Research Infrastructure for High Performance Computing (HPC) - <a href="http://www.prace-project.eu/">http://www.prace-project.eu/</a></td>
</tr>
<tr>
<td><strong>PTB</strong></td>
<td>Project Technical Board</td>
</tr>
<tr>
<td><strong>QA</strong></td>
<td>Quality Assurance</td>
</tr>
<tr>
<td><strong>ROC</strong></td>
<td>Regional Operation Centre</td>
</tr>
<tr>
<td><strong>S2I2</strong></td>
<td>Scientific software Innovation Institute</td>
</tr>
<tr>
<td><strong>StratusLab</strong></td>
<td>Enhancing Grid Infrastructures with Virtualization and Cloud Technologies (<a href="http://stratuslab.eu/doku.php?id=start">http://stratuslab.eu/doku.php?id=start</a>)</td>
</tr>
<tr>
<td><strong>TCB</strong></td>
<td>Technology Collaboration Board</td>
</tr>
<tr>
<td><strong>TERENA</strong></td>
<td>The Trans-European Research and Education Networking Association (<a href="http://www.terena.org/">http://www.terena.org/</a>)</td>
</tr>
<tr>
<td><strong>UMD</strong></td>
<td>Unified Middleware Distribution, the EGI middleware distribution</td>
</tr>
<tr>
<td><strong>UNICORE</strong></td>
<td>The Uniform Interface to Computing Resources offers a ready-to-run Grid system including client and server software. UNICORE makes distributed computing and data resources available in a seamless and secure way in intranets and the internet (<a href="http://www.unicore.eu/">http://www.unicore.eu/</a>)</td>
</tr>
<tr>
<td><strong>VENUS-C</strong></td>
<td>VENUS-C draws its strength from a joint co-operation bringing together industrial partners and scientific user communities to develop, test and deploy an industry-quality Cloud Computing service for Europe (<a href="http://www.venus-c.eu/Pages/Home.aspx">http://www.venus-c.eu/Pages/Home.aspx</a>)</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>VRC</strong></td>
<td>Virtual Research Community</td>
</tr>
<tr>
<td><strong>WLCG</strong></td>
<td>The Worldwide LHC Computing Grid (WLCG) is a global collaboration of more than 140 computing centres in 34 countries, the 4 LHC experiments, and several national and international grid projects (<a href="http://lcg.web.cern.ch/lcg/">http://lcg.web.cern.ch/lcg/</a>)</td>
</tr>
<tr>
<td><strong>WP</strong></td>
<td>Work Package</td>
</tr>
</tbody>
</table>
PART 1

THE EMI DEVELOPMENT ROADMAP AND ITS STRATEGIC POSITIONING IN THE CONTEXT OF EUROPEAN AND INTERNATIONAL COLLABORATIONS
1 EXECUTIVE SUMMARY

The document is organized in two main parts. Part 1 provides the description of the overall EMI strategy linking the high-level EMI technical objectives with the software release roadmap and the collaboration activities with other projects and initiatives in Europe and internationally.

The high-level overview of the EMI Development Roadmap is presented in section 1-2. It outlines outlining the relationships and synergies across the various EMI activities and provides an overall timeline of the major phases.

Existing or planned collaboration activities with projects from the European Distributed Collaboration Infrastructure (DCI) projects and from international projects from the US and Asia-Pacific areas are also described in sections 1-3 and 1-4 to show how EMI relates to important international initiatives and proposes to contribute to the global international infrastructure development and standardization.

EMI is also committed to finding ways of collaborating with and engaging commercial partners into exploiting the EMI middleware services and the distributed infrastructures to provide commercial services. The services provide can range from the provision of direct technical support for existing middleware applications to the provision of consultancy or development of applications for targeted scientific and industrial domains. The EMI strategy in this area is described in section 1-5.

The second part of this document is the “DCI Collaboration Roadmap” document, a unified vision and collaboration roadmap defined in close partnership by the six DCI projects co-funded by the European Commission as part of INFRA-2010-1.2.1: Distributed computing infrastructure (DCI). This document describes in more details the vision and strategy of the specific collaboration among the six projects. It also presents a description of the opportunities for collaboration among the projects in group or as bilateral activities. The activities will be formalized by signed MoUs during the course of the projects. The document presents the overall strategic vision for EMI and the other DCI projects and is an integral part of the EMI Roadmap document. By agreement of the six projects and with the endorsement of the European Commission, it is provided without modifications with the exception of the formatting required to fit it into the EMI document template.
2 THE EMI ROADMAP

2.1 INTRODUCTION

The European Middleware Initiative is a close collaboration of four major European middleware providers, ARC, gLite, UNICORE and dCache. Its main objectives are to deliver a consolidated set of middleware components for deployment in EGI (as part of the Unified Middleware Distribution or UMD), PRACE and other DCIs, extend the interoperability and integration between grids and other computing infrastructures, strengthen the reliability and manageability of the services and establish a sustainable model to support, harmonise and evolve the middleware, ensuring it responds effectively to the requirements of the scientific communities relying on it.

EMI has a specific focus on providing high-quality middleware solutions to well-defined user requirements collected, analysed, prioritized and implemented using a constructive network of collaborations and relationships with other project and initiatives sharing common interests. The goal of these collaborations is to define common objectives, perform joint activities, and reuse knowledge and technology produced by either party to optimize effort and speed up the implementation of the objectives. The main targets (Customers) of the EMI activities are the major European and international infrastructure projects, scientific user communities developing applications to run on the infrastructures and making use of the middleware services, standardization bodies and commercial companies offering distributed computing services to their customers.

In order for the EMI customers to have clear communication channels with EMI, be well informed of EMI products and services and plan for their own activities, EMI has defined a concrete roadmap of linked activities, which include participation to strategic and technical decisional bodies within other projects and initiatives, collaboration programs, planned software design, development and release cycles, user support and training activities and comprehensive dissemination and use of knowledge activities. The details of the collaboration programs, the dissemination and training plans, the sustainability plans and the user support activities are described in specific deliverables from the NA2 [R1, R2, R3, R4] and SA1 [R5] Work Packages. This document serves as high-level introduction to the more specific project deliverables and provides an overview of the EMI roadmap, highlighting the major objectives and milestones and putting the more detailed plans into the general context of the EMI overall strategy.

The EMI project aims at addressing and solving a number of problems that today still prevent users from easily accessing and using the existing computing infrastructures, at helping infrastructure projects to decrease the cost of operations and ownership, and at providing a robust and open channel for users to provide new requirements and follow the status of their implementation in the middleware.

In order to achieve this mission, EMI relies on a number of high-level technical objectives targeted at addressing a well defined set of requirement categories as described in the EMI Project DoW [R6] and in the DNA1.3.1 – Technical Development Plan deliverable [R7]. In particular EMI needs to work on the following areas:

- **Usability** must be enhanced by removing redundancy and consolidating the services, simplifying the security management without compromising its strength, adding better programmability interfaces and support for developing user gateways and portals and transparently making use of virtualization to increase resource availability and management.

- **Interoperability and compatibility** must be improved by removing proprietary interfaces in the middleware services and ensuring true interoperability through the adoption of community
standards whenever possible or at least providing uniform interfaces across the four middleware distributions represented in EMI. Interoperability between grids, supercomputers and emerging computing models like clouds and desktop grids must be extended to address scalability and accessibility requirements.

- **Manageability** of the services must be improved by providing standard service configuration mechanisms, monitoring and instrumentation interfaces and making accounting and other operational information more readily accessible.

- **Sustainability** must be improved by establishing collaboration programs with commercial companies, by better integrating the middleware components in the established open source operating system distributions and by adopting off-the-shelf components whenever possible to reduce maintenance costs and to facilitate easier adoption by wider user communities.

The initial set of requirements has been inherited by EMI from past or present projects like EGEE, KnowARC, NorduGrid, DEISA, OSG and others and further refined during the project proposal preparation through extensive contacts with user communities and other project proposals coordinators, in particular EGI-InSPIRE and WLCG.

### 2.2 THE EMI STRATEGIC VISION

The roadmap to achieve the EMI vision for the next three to five years has been initially described in the project Description of Work and further elaborated in collaboration with EGI-InSPIRE and the other DCI projects as part of the DCI Collaboration Roadmap provided in Part 2 document, which represents an important example of collaboration, where the six projects together have joined forces to create a single vision all of them can contribute to, rather than defining six separate visions.

The EMI vision foresees that after the necessary initial period of research and consolidation that took place in the past 6 to 8 years, the growing usage of distributed computing and data resources by scientific communities and individual researchers requires now the stabilization of the computing infrastructures and a simplification and standardization in the use of the associated software tools. It is of strategic importance towards the establishment of permanent, sustainable research infrastructures to lower the barriers that still prevent potential communities of tens of thousands of scientists and researchers to consider grids as a commodity tool serving their daily research activities. The software must allow for easy extension and integration of existing and new technologies, like virtualization and on-demand provisioning methods, and the middleware must be largely based on open standards to allow for the creation of a marketplace where competition in the provision of added-value services can take place.

The ultimate vision is that establishing distributed scientific collaborations and using distributed computing and data resources should be as easy as opening a web application, entering simple identification information, entering a few clear parameters to define the task to be executed and its requirements and then waiting for the results to be made available in a well known, easily accessible place.

### 2.3 THE EMI ROADMAP AT A GLANCE

The EMI strategic vision is implemented through its development and collaboration roadmap. The roadmap elements and objectives are described in the following sections 2.3 and 2.4. This section gives a synoptic view of the major milestones of the roadmap.
<table>
<thead>
<tr>
<th>Milestone</th>
<th>Impact</th>
<th>Expected date</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Software Maintenance and Support Plan is available</td>
<td>This plan is the starting point for every collaboration, since it describes what EMI provides to its users and customers</td>
<td>PM 1 (31 May 2010)</td>
</tr>
<tr>
<td>EMI 1 Technical Plans are available</td>
<td>The technical plans are the starting point of the yearly release cycle</td>
<td>PM 3 (31 July 2010)</td>
</tr>
<tr>
<td>MoU signed with EGI</td>
<td>Formalise the relationship with EGI on shared objectives</td>
<td>PM 6 (30 October 2010)</td>
</tr>
<tr>
<td>Relationships with OSG is established</td>
<td>Positions EMI and its relationship with the main US infrastructure project</td>
<td>PM 6 (30 October 2010)</td>
</tr>
<tr>
<td>SLA signed with EGI</td>
<td>Formalise the relationship with EGI on service provision</td>
<td>PM 10 (28 February 2011)</td>
</tr>
<tr>
<td>MoUs signed with other DCI projects</td>
<td>Formalise the relationship with the DCI projects on shared objectives</td>
<td>Between PM10 and PM 12 (February to April 2011)</td>
</tr>
<tr>
<td>First MoUs with commercial companies</td>
<td>First definition of possible shared objectives with commercial companies to be used as prototype for further collaborations</td>
<td>PM 12 (April 2011)</td>
</tr>
<tr>
<td>Long-term sustainability strategy is defined</td>
<td>Necessary starting point of the long-term sustainability roadmap</td>
<td>PM 12 (April 2011)</td>
</tr>
<tr>
<td>EMI 1 is released</td>
<td>First major release of the EMI distribution, integration of the four middleware providers services, changes required to add standard OS distribution compliance as the first step of the sustainability plan</td>
<td>PM 12 (April 2011)</td>
</tr>
<tr>
<td>EMI 2 Technical Plans are available</td>
<td>The technical plans are the starting point of the yearly release cycle</td>
<td>PM 15 (31 July 2011)</td>
</tr>
<tr>
<td>Relationship with Latin America and Asia-Pacific is established</td>
<td>Positions EMI and its relationship with the main Latin Americas and Asia-Pacific infrastructure project</td>
<td>PM 18 (31 October 2011)</td>
</tr>
<tr>
<td>Additional MoUs with commercial companies</td>
<td>Exploration of what services and how can be outsourced to commercial companies in the terms of support and further</td>
<td>PM 24 (30 April 2012)</td>
</tr>
<tr>
<td>Event Description</td>
<td>Description</td>
<td>Relevant PMs</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>MoU signed with PRACE</td>
<td>Formalise the relationship with PRACE on shared objectives</td>
<td>PM 24 (30 April 2012)</td>
</tr>
<tr>
<td>SLA signed with PRACE</td>
<td>Formalise the relationship with PRACE on service provision</td>
<td>PM 24 (30 April 2012)</td>
</tr>
<tr>
<td>EMI 2 is released</td>
<td>Second major release of the EMI distribution, major functional changes as per DoW and new user requirements, especially security, harmonization, standardization, messaging, etc</td>
<td>PM 24 (30 April 2012)</td>
</tr>
<tr>
<td>EMI 3 Technical Plans are available</td>
<td>The technical plans are the starting point of the yearly release cycle</td>
<td>PM 27 (31 July 2012)</td>
</tr>
<tr>
<td>Full compliance with relevant open standards</td>
<td>Compliance with open standards is an important element of the ease of use improvements and of the sustainability plan implementation</td>
<td>PM 20 (31 December 2012)</td>
</tr>
<tr>
<td>Most EMI Services and Clients are in Fedora/EPEL, Debian or Ubuntu</td>
<td>Integration with mainstream OSs is an important element of the ease of use improvements and of the sustainability plan implementation</td>
<td>PM 20 (31 December 2012)</td>
</tr>
<tr>
<td>Additional MoUs with commercial companies</td>
<td>Additional outsourcing activities to commercial companies in the terms of support and further development. By now the market analysis must be completed and there should be a better understanding of how support and development services can be provided by commercial companies</td>
<td>PM 20 (31 December 2012)</td>
</tr>
<tr>
<td>Long-term sustainability strategy is implemented</td>
<td>The strategy defined at the end of the first year should be implemented by now. All the elements of simplification, standardization, integration with standard OSs and definition of the market and the business model should be in place</td>
<td>PM 20 (31 December 2012)</td>
</tr>
<tr>
<td>EMI 3 is released</td>
<td>Third and final major release of the EMI distribution as “EMI Reference Services” fully</td>
<td>PM 22 (28 February 2013)</td>
</tr>
</tbody>
</table>
compliant with the long-term sustainability and technical strategies

| Table 1: The EMI Roadmap at a glance |

2.4 THE EMI SOFTWARE DEVELOPMENT ROADMAP

In order to implement the vision EMI requires a clear definition of how its software development strategy will evolve during the three years of its duration and what are the major expected objectives.

The EMI Software Development Roadmap is represented in Figure 1. The roadmap consists of three major phases:

1) At the start of EMI the four major European middleware providers, namely ARC, gLite, UNICORE and dCache join effort in EMI committing to work together on the achievement of the project objectives. They bring their software products and existing set of requirements.

2) During the three year duration of the EMI project, the EMI developers and engineers work together to consolidate, harmonise and support the existing software products, evolving and extending them based on existing and new requirements. Redundant or duplicate services resulting from the merge are deprecated; new services can be added to satisfy user requirements or specific consolidation needs. Input for the development activities is taken from users, infrastructures projects, standardization initiatives or evolving technologies. The software components are adapted as necessary to comply with standard open source guidelines to favour the integration in mainstream operating system distributions [R7].

3) At the end of EMI, a robust set of Reference Services emerges. The software components have standard interfaces, are well integrated in the operating systems and distributed through standard channels. Support for the services can be provided by the original Middleware Collaborations or commercial companies using standard SLA-based contracts. The Reference Services can be further extended through typical open source models or by professional application providers.

The overall EMI Roadmap is structured on four major releases, called EMI 0 (Zugspitze), EMI 1 (Kebnekaise), EMI 2 (Matterhorn) and EMI 3 (Monte Bianco). The sequence of releases is shown in Figure 2.

The releases are time-based. They are scheduled in advance for a certain date and a proposed content described in the Development Plans. As the release date approaches, components that are found not to meet the release criteria in time for the release are dropped and rescheduled for another release.

The release called EMI 0 is an internal release meant to be used as internal benchmark for validating the software engineering processes, the actual skills of EMI in managing the release, and specifically to construct a solid baseline for the public releases after solving the expected integration issues coming from merging together products from four middleware collaborations (ARC, gLite, UNICORE and dCache). The EMI 0 release follows an abbreviated cycle, where essentially only a subset of the integration and deployment testing is done as part of the Certification phase. Apart from EMI 0, all other releases follow the standard cycle and are followed by a period of support and maintenance according to the policies defined in the Software Maintenance and Support Plan [R5]. How to provide
support for EMI 2 and EMI 3 after the end of EMI is one of the important questions to be addressed by the EMI Sustainability Plan [R4].

The major releases are checkpoints used to make sure that all technical objectives planned for that release are indeed implemented and working as expected, that all interface contracts among components are respected and that the acceptance criteria defined by the customers are all satisfied for all released components.

Between any two releases, each EMI Product can follow an internal release cycle similar to the overall release cycle, but with shorter timelines. The only constraint that each Product has to respect, apart from the overall acceptance criteria, is that its interfaces have to be backward compatible with the release currently in production. Any non backward compatible change must be released only as part of a major EMI Release at the end of the main release cycle. Using this release cycle, specific products versions can be released at any time during the year as minor releases or bug fix releases, following the plan defined in the development and release plans at the very beginning of the overall release cycle.

**Figure 1: The EMI software Development Roadmap**
The major EMI releases must fulfill specific overall goals within which the specific technical objectives are defined every year based on user requirements. The overall goals are:

**EMI 1 (Kebnekaise)**: the first release of the EMI distribution has the overall goal of providing the first integrated set of components and services from all four participating middleware distributions. In addition, it must introduce formal compliance with the most important software release policies from major operating system distributions. The level of new functionality in this release is not expected to be significant, however the majority of components will have to be modified in their structure, deployment models, and level of integration with the operating systems. Most components are expected not to be backward compatible in their deployment, although they will be backward compatible in terms of functionality. These changes represent a necessary step towards the goal of better integration with standard OSs and improvements of the ease of use of the middleware stack. It is also a necessary step to be able to expand the support for additional platforms toward the current limited offer of the majority of the components.

**EMI 2 (Matterhorne)**: the second release of EMI capitalizes on the structural changes introduced in EMI 1 to introduce an important number of new and improved functionalities and expand support for more platforms. The functionality introduced in this release is expected to be in most part as described in the EMI DoW. However, at the end of the first year the user requirements will be re-evaluated and the necessary changes in the technical plans will be introduced. This release is also expected to introduce a number of important changes in the adoption of common interfaces across the four middleware stacks and better compliance with open standards.

**EMI 3 (Monte Bianco)**: the third release of EMI will provide the so-called “EMI Reference Services”. It will necessarily contain functional improvements according to the requirements described in the DoW or new and changed requirements from the user communities. However, it will above all
conclude the implementation of the simplification and standardization objectives of the project and provide full integration within standard operating systems. The goal is that by the time EMI 3 is released, most major EMI services and clients are directly available to the user communities as part of the major OS distributions like Fedora/EPEL, Debian and Ubuntu with certified compliance with all relevant open standards. This is an important objective in terms of the EMI sustainability plan, since it is believed that only such a level of integration with the standard open source software ecosystem can expand the user base while at the same time decreasing the development and maintenance costs.

2.4.1 The yearly development cycle

The maintenance and development of the EMI middleware services is based on a 5-step yearly cycle. The cycle is represented in Figure 3 and is composed of:

1) **The Requirements Analysis phase:** input collected from the EMI collaboration activities or from direct user submission in the EMI User Support system is analysed and prioritized based on the Severity assigned by the users, the urgency, impact, cost and available effort. The result of the analysis is compiled in the form of the EMI Technical Plans defining the project technical objectives (EMI Deliverables DNA1.3.x, DJRA1.1.x, DJRA1.2.x, DJRA1.3.x, DJRA1.4.x, DJRA1.5.x and DJRA1.6.x). The plans are defined at the beginning of the project and refined at every cycle based on the new input. The requirements analysis and the overall technical plan are coordinated by the PTB and the specific Technical Areas and Standardization Plan are coordinated by the Technical Area leaders and the Standardization Task leader within JRA1.

2) **The Development and Test Plans phase:** based on the latest version of the project Technical Plans, the Development and Test Plans DJRA1.8.x for the current cycle are defined. The plans outline which of the technical objectives can be included in the cycle, which components are involved, which platforms and operating systems can be targeted, external and internal integration constraints, development, deployment and testing timelines, etc. The plans are centrally coordinated by the JRA1 Work Package leader and distributed to the EMI Product Teams for implementation. The Release Plan is established by SA1 [R8] with the details of the timelines to be applied to each release cycle (code freeze date, release date, any technical preview release date) and the outline of the set of acceptance criteria to be fulfilled by the components (documentation, specific categories of tests, etc).

---

1 The set of deliverables DJRA1.8.x were not foreseen in the DoW. However, given the importance of these documents also as input to the EGI UMD preparation plans, it is proposed to add them and present them officially at the EC reviews.
3) **The Development, Testing, Certification phase**: based on the Development and Test Plans, the various Product Teams develop the new functionality and perform initial unit, integration, deployment and functional tests under the overall coordination of JRA1 and monitoring of the PTB. Once a piece of functionality has passed the foreseen set of acceptance criteria and has been certified, it is released as Release Candidate to SA1.

4) **The Release Certification and Validation phase**: as Release Candidates are transitioned from JRA1 to SA1, the final certification phase starts. During this phase, the software components are validated against the set of acceptance criteria defined by the Customers, Technical Previews are made available to users and projects taking part in the “Works with EMI” technical collaboration program [R1] and the final packaging and signing is performed. Components who do not pass the criteria are rejected back to JRA1 for revision.

5) **The Release and Maintenance phase**: once the software components have passed all acceptance criteria, they are released and uploaded to the official EMI Software Repository from where they can be picked up by users and infrastructure operators. The continuous maintenance phase starts at this point, any defect found by users in production environments and submitted to the EMI User Support system (GGUS) are analysed, prioritised and addressed as revision or minor releases. Any component that has not passed all criteria by the time the final release is due it is rejected and reschedule for another release cycle.

During the whole cycle, SA2 is responsible to monitor the actual application of the Software Engineering guidelines as defined in the Software Quality Assurance Plan [R9] and periodically reported in the deliverables DSA2.3.x.
2.5 THE EMI COLLABORATION ROADMAP

Most of the input collected by EMI to prepare the technical plans and the release cycles comes from the network of collaborations established by EMI very early in the project. Details of the different collaboration programs and partnerships models are given in the NA2 Collaboration Programs deliverable (DNA2.1.1) [R1].

In this section a high-level view of the existing collaborations is presented in the overall context of the EMI Roadmap development.

EMI has identified four major targets for collaborations, as shown in Figure 4.

The primary target is represented by major European and non-European infrastructure providers, namely EGI, PRACE, WLCG, OSG and similar initiatives. Input from these initiatives is collected through a continuous collaboration and in particular at the beginning of every new major release cycle. Conversely, EMI provides software releases and technical support, as much as possible based on formal Service Level Agreements or at least Memorandums of Understanding.

The second target is represented by the user communities or more specifically by technical developers or software engineers within user communities developing domain-specific tools, like scientific gateways and portals. In most cases requirements for user communities can also come from equivalent user-oriented activities in EGI, PRACE and others, but EMI maintains specific technical relationships during the later stages of the process, when interfaces or programming libraries have to be discussed and presented directly to the technical experts. The relationships with the communities are maintained through the mechanisms described in DNA2.1 and by means of dissemination and collaboration events, like conferences, workshops, etc.

Figure 4: The EMI Collaborations Map
The third target is represented by industrial partners and standardization bodies. The two categories are handled together since most of the commercial partners have been found to be more likely to interact with mostlyacademic projects like EMI only if the software specifications are well documented and widely used. Also in this case, EMI has a specific collaboration program for industrial partners and is very actively participating to international standardization activities. The goal is to converge in the long terms to widely accepted open standards in the respect of the stringent, short-term objectives of providing today efficient solutions to user problems at least within Europe and the EGI community using the EMI products. More details on the EMI Standardization plans are given in the JRA1 deliverables DJRA1.5.x.

The fourth target is represented by the five DCI projects co-funded by the EC in the call INFRA-2010-1.2.1: Distributed computing infrastructure (DCI) together with the SIENA project. The seven projects together (including EMI) have established a close collaboration and defined a development roadmap to be carried out together during the lifetime of these projects. The goal is to create a self-sustainable infrastructure reducing the need for EC funding for maintenance, support and operations, increasing the participation of local national agencies and industrial partners, while introducing innovation and new technology in a controlled and non-disruptive way.

The strategic positioning of EMI within the DCI projects, with projects outside Europe and with industrial partners is described in the following sections 3 to 5. The main collaboration programs and mechanisms are described in the NA2 Collaboration Programs deliverable (DNA2.1.1) [R1]. In summary, EMI will establish MoUs and SLAs with all projects and initiatives sharing common objectives. The MoU and the SLA are considered to fulfil two different and complementary roles:

**MoU**: Memorandums of Understanding will be established with all projects and initiatives with which EMI can work on the achievement of common objectives. The MoU describes the objectives and a set of associated tasks, milestones and responsible persons. This is considered to be the first level of collaboration and it is open to all projects and initiatives.

**SLA**: Service Level Agreements are a more formal description of the levels of service that specific EMI customers can expect. They apply to the range of services provided by EMI as described by the EMI Software Maintenance and Support Plan [R5]. SLAs are established with selected projects and initiatives on the base on the establishment of a common long-term business strategy. They are formally signed and monitored.

SLAs have themselves a dedicated roadmap to follow as described in DN1.2.1 [R10]. The first instance of the EMI SLA is based on ITIL guidelines and contains a formal description of the agreed service levels, but no financial commitments or constraints. This is due to the fact that normally both parties in the SLA, EMI and its customer, do not enter into financial contractual agreements at this stage, but are funded by the EC and their respective project partners. However, the SLA is revised every year and it is expected that a number of services and the associated service levels will transition to a more commercial model with contractually formal financial constraints. This is also an important aspect of the EMI sustainability plan.
3 THE DCI PROJECTS AND OTHER EUROPEAN COLLABORATIONS

Call 7 (FP7-Infrastructures-2010-2) under the e-Infrastructures topic of the FP7 "Capacities" Specific Programme which closed in November 2009 called for proposals under the topic of ‘Distributed Computing Infrastructures’. As a result of this call, six projects were funded that together are referred to in this and related documents as the ‘Distributed Computing Infrastructure Projects’ or ‘DCI Projects’. Each of the six projects is described in more details in Part 2, Annex A.

Together the projects fund different activities in the area of distributed computing infrastructures ranging from the provision of production environments, the development, maintenance and support of the middleware used in Europe, and the exploration of the provision and use of virtualised computing resources. Due to the competitive nature of the funding model, many of these projects were developed in isolation in order to maintain confidentiality during the proposal phase. As a result, one of the goals required by the European Commission is that at the start of all of these projects is for each project to establish how they plan to collaborate with each other (if at all) and what the results of those collaborations may be within the scope of the project, and the impact that there may be long-term within the community.

The common collaboration program established among the six DCI Projects is described in details in the common roadmap document, developed together by the six project and presented at the EGI Technical Forum in September 2010, which is reported in integral form in Part 2.

This section provides more specific details of how this on-going collaboration is being implemented.

3.1 EGI-INSPiRE

The EGI-INSPiRE project is the major customer and source of requirement for EMI. The two projects have closely collaborated since the initial phases of the proposal preparation activities and are now formalizing their relationship with concrete common initiatives.

The EMI and EGI common objectives are described in more details in Part 2, section 4.2.1. Those objectives are implemented through a number of formal mechanisms:

1) **Requirements analysis and technical roadmaps**: the EGI-INSPiRE project has established in October 2010 the Technology Collaboration Board (TCB). The TCB is composed of the leader of the major EGI activities (user communities, operations, policies and software auditing) and representatives of the software providers developing and supporting software for EGI, among which are EMI and IGE. This board is responsible to discuss high level requirements coming from user communities or EGI activities and get preliminary technical information on how the requirements can be incorporated into the EGI UMD Roadmap and the specific technical roadmaps of EMI and IGE. Once the requirements are formalised by the TCB, they can be further discussed by EMI directly with technical experts in EGI or the relevant user communities. This formal channel is now being complemented by the establishment of a Memorandum of Understanding and a Service Level Agreement, which are foreseen to enter into force in December 2010.

2) **Software releases**: EMI software is released to EGI using the software release process described in section 2.2. The specific mechanism to announce releases and validate the acceptance criteria have been defined together with EGI and are implemented by the EMI SA1 Work Package and the EGI WP5 Work Package (Provisioning the Software Infrastructure). The tasks and milestones are formalized by a Memorandum of Understanding to be signed by EMI and EGI in December 2010.
3) **User Support:** EMI and EGI are part of an overall distributed user support infrastructure designed to optimize the response times and the load on support personnel and technical experts. In this structure EGI provides 1st and 2nd level support (via the local national help desks and the EGI DMSU respectively), while EMI provide 3rd level expert support for the middleware services under its responsibility. This model allows escalating incidents requiring modifications in the software to the appropriate technical experts without burdening them with incidents that can be solved using know workarounds, documentation clarifications, standard configuration changes, etc. The EMI User Support policies are described in deliverable DSA1.1 [R4], the agreed service levels are described in a Service Level Agreement between EMI and EGI to be signed in February 2011 and based on the EMI Service Level Agreement Template (DNA1.2.1) [R10].

### 3.2 OTHER DCI PROJECTS

EMI is also implementing dedicated channels to each of the other DCI projects to make sure that the collaboration takes the form of a continuous activity with well defined milestones. In particular, EMI is establishing MoUs with IGE, EDGI, StratusLab and VENUS-C and has started its official collaboration with SIENA by nominating an EMI representative in the SIENA standardization activities.

In the case of IGE, EDGI, StratusLab and VENUS-C, contact points for each project have been nominated on both sides. All projects have therefore established dedicated channels were the actions and milestones described in the MoUs will be monitored and discussed. In addition, a number of technical task forces have been created within JRA1 to investigate specific technological topics, which are of interest not only to EMI, but also by the other projects. Task forces on virtualization, cloud computing and accounting, are for example working in close collaboration with members of StratusLab and VENUS-C to investigate common problems and possible solutions.

### 3.3 PRACE AND OTHER HIGH-PERFORMANCE COMPUTING INITIATIVES

The PRACE project is another important source of requirement for EMI in particular in the domain of High Performance Computing. Members of the project preparation teams of EMI and PRACE have collaborated since the initial phases of the proposal preparation activities and are now formalizing their relationship. Since the 1st Implementation Phase of the PRACE project officially started in July 2010, it was not possible to establish formal contacts until late September 2010.

In November 2010 EMI and PRACE have formally agreed that PRACE technical representative will participate to the review process of the EMI Technical Plans. Since PRACE does not have yet an official body for discussing requirements for the middleware services, EMI will interact with PRACE by attending the major PRACE events to discuss status and requirements directly with the relevant technical experts. If a more formal collaboration becomes necessary as PRACE progresses in its implementation work plan, the collaboration model will be revised.

EMI is also planning to establish collaboration relationships with other HPC initiatives in Europe and internationally in its second year of activities. In particular, EMI plans to collaborate with HPC-SEE to understand how their infrastructure places itself in relation to EGI and PRACE and possibly collect and implement specific requirements in the middleware. High-level discussions with TeraGrid have also started and similar discussions will take place with the Japanese RENKI project. Since TeraGrid is basing its Science Gateways on top of the Globus middleware and RENKI is developing their own middleware, the major objective in this additional collaboration with them is to discuss and achieve interoperability of at least the most important services.
3.4 WLCG
The World-wide LHC Computing Grid is by far the largest and most active user community using the EMI middleware services. WLCG has a well-defined structure for managing and monitoring the resource infrastructure, the deployed services and the development of software. Although a large part of the infrastructure and services used by WLCG are managed by sites participating in EGI, WLCG has expressed the need to directly control the process of defining new requirements and implementing them in future middleware services. To this end, WLCG is establishing a technical board that will collect and analyse the requirements of the HEP community and make recommendations on how the middleware services (and also other software) have to evolve. In case any of these requirements has to be implemented in services supported by EMI, they will be passed to EMI via the WLCG Management Board, a high-level collaboration board composed of representatives of the WLCG stakeholders. As of September 2010, EMI is an official member of the WLCG MB and in this role will periodically report on the status of the EMI middleware of interest for WLCG and receive input about requirements.

3.5 EEF
The European EInfrastructures Forum is a high-level collaboration initiative among EGI, DEISA, PRACE, Terena and GEANT to collect requirements, identify common needs and promote common initiatives with particular focus on supporting the ESFRI projects and other virtual research communities. Although EMI has already direct contacts with at least EGI and PRACE, the EEF provides invaluable input on common requirements usually having rather high priority in the work plans of the EEF members. EMI has therefore established links with EEF, providing an initial mapping of its planned features against the already defined set of EEF requirements and including in its work plan additional requirements. EMI has presented the current status of the implementation at the EEF meeting that took place at CERN on November 2nd, 2010¹ and will periodically report to EEF and collect additional input.

3.6 ESFRI PROJECTS
The European Strategy Forum on Research Infrastructures is a forum tasked with supporting the definition of strategies for the development of the research roadmap in Europe and facilitating the use of Research Infrastructures. A number of projects in various research domains have been and will be funded to implement the roadmap. The projects make heavy use of the resource infrastructures and the middleware services, often developing domain-focused applications. They are therefore an extremely important source of requirements for EMI, especially in critical areas like security, interoperability, standardization, usability, etc. EMI receives relevant requirements from the existing ESFRI collaboration with EGI, PRACE and the EEF. However, once a requirement has been identified and planned for, the more technical aspects of its design and implementation needs to be discussed directly with domain experts from individual projects. EMI has a specific collaboration program for ESFRI projects described in more detail in the EMI Collaboration Programs deliverable [R1].

3.7 STANDARDIZATION BODIES

EMI has a strategic interest in participating in the design, development and implementation of international open standards. The goal is to simplify the service design, improve interoperability and allow further development by third-parties. The collaboration with standardization bodies like OGF or OASIS is therefore a key point in the EMI development roadmap. However, it is also important to stress the fact that EMI brings together middleware services from different providers with existing non-standard interfaces. The primary goal of EMI is to ensure that all service within the EMI stack adopt common interfaces. If suitable open standard interfaces exist, EMI is committed to adopt them. If they do not exist, EMI needs first of all to design and adopt common interfaces across the stack.

Once the EMI interfaces have been normalized and the most immediate interoperability issues within EGI and between EGI and other infrastructures are addressed, EMI is committed to disseminate the results of the harmonization work into international standardization bodies and start an open discussion and standardization process for the most relevant results. The initial EMI standardization plan is described in details in DJRA1.5.1 [R11].
4 OSG AND OTHER INTERNATIONAL COLLABORATIONS

4.1 OPEN SCIENCE GRID

The Open Science Grid has a long story of collaboration with European infrastructure projects especially with EGEE. During the preparation phase of the EMI project, contacts were established with OSG representatives to collect feedback and make sure that the collaboration would continue after the transition from EGEE to EGI and EMI. OSG has signed a Letter of Support and Commitment that was attached to the EMI proposal and close relationships have since then been established.

OSG and EMI are collaborating especially in two areas:

1) **Technical collaboration**: from the technical point of view OSG and EMI have common interest in keeping the software and the infrastructures interoperable and to converge to a common adoption of open standards via collaboration with the appropriate organizations like OGF. Traditionally, middleware produced in Europe by EGEE or KnowARC has made used of components developed or supported by US initiatives like the Globus project or the VDT project, which provide middleware service to OSG in a very similar way to what EMI does for EGI. Conversely a number of European middleware services like the VOMS authorization service or the CREAM Compute Element service are used by OSG sites and are included in VDT releases, but may in the future be taken by OSG directly from EMI or from standard OS distributions where EMI services are integrated. This collaboration will keep running during EMI with an even stronger focus on interoperability and usability. OSG technical experts have been invited to review the EMI technical plans and EMI experts are members of the ongoing OSG WLCG Interoperability working group.

2) **Strategy collaboration**: as in Europe also in the US the discussion on how to make the research infrastructures and the software development sustainable is the focus on most of the strategic funding initiatives. Recently the National Science Foundation (NSF) has issued a request for recommendations about the establishment of a Scientific Software Innovation Institutes (S2I2). EGI and EMI are already trying to solve some of the issues that the S2I2 should address and it is acknowledge that close collaboration on establishing sustainable and long-term coordination activities is necessary. EMI has been invited to take part into a set of workshops dedicated to collecting feedback on the roles and tasks of a future S2I2. The collaboration is planned to continue in the future possibly with the definition of common proposals for joint funding from the EC and NSF.

4.2 LATIN AMERICA COLLABORATIONS

The collaboration with infrastructures and communities in the Latin America Countries has been traditionally organized by means of dedicated projects like EELA and GISELA or through the participation of relevant resource sites in the EGEE and WLCG Regional Operations Centers (ROCs) coordination meetings. EMI aims to provide middleware and support to Latin America infrastructures either through their relationships with EGI or with direct collaboration initiatives. The first of such initiative has been tentatively agreed during the 1st Workshop of the Latin America ROC that took place at CERN in October 2010. During this meeting it has been agreed that a foreseen new resource

---

3 https://sites.google.com/site/distributedcomputings2i2/home
infrastructure to be set up in Brazil during the first half of 2011 could provide middleware services from the EMI 1 distribution possible also acting as certification site before the official release.

4.3 ASIA-PACIFIC COLLABORATIONS

EMI has official project partners from the Asia-Pacific area, since traditionally the collaboration between the European infrastructure projects and equivalent projects in Asia has been very close. The presence of Academia Sinica from Taiwan and KISTI from South Korea ensure the continuation of the collaboration in EMI in several important middleware development areas, like metadata catalogues, and integration between Grids and Clouds. In addition, EMI is establishing direct links with the major Europe-Asia collaboration projects. Memorandums of Understanding with CHAIN and EU-IndiaGrid are being discussed. The plan is to work on common requirement analysis, application porting and training activities. This part of the EMI roadmap will mostly start in the second year of the project and will be further refined after the release of the first EMI major release.
5 INDUSTRIAL COLLABORATIONS

The establishment of collaboration initiatives with industrial partners is an important element of the EMI sustainability strategy. In previous infrastructure and middleware projects a number of tentative collaborations have been established, but there are few examples of actual persistent adoption of grid technology by commercial and industrial companies as the result of transfers from academic projects or joint efforts. One of the objectives that EMI plans to achieve is to better understand the business model of grid services provision and actively address as realistically as possible the constraints that have prevented a better transfer of grid-based distributed computing technologies and competencies from EC-funded projects to commercial activities. More details on the industrial collaboration programs are given in the EMI Collaborations Programs deliverable DNA.1 [R1].

EMI plans to achieve this objective by working with interested industrial partners on two specific fronts:

1) **Analysis of the business model**: the first step towards a convincing commitment of industrial partners in engaging with grid projects and later directly developing, using or supporting grid technologies is the understanding of the possible business models, the target markets and the possible revenue streams. Such analysis requires specific competences that are extremely difficult to be found in academic Institutes. EMI is therefore trying to engage large and small commercial companies in providing the necessary knowledge of the methods and the markets to and come to such a definition at the end of the second year of EMI. It is understood that the result of the analysis is open, in the sense that it may provide a definition of possible viable business models, but may also come to the conclusion that the use of distributed grid infrastructures based on the concept of resource sharing may not be suitable for commercial companies, despite their success in the research environment.

2) **Technology transfer**: the second collaboration front concerns the mutual exchange of ideas and technologies between EMI partners and commercial companies. The exchange is meant to enrich the EMI middleware with specific requirements or third-party off-the-shelf products and transfer technologies, knowledge and support services from EMI to the companies.

Both activities rely on the progressive simplification and standardization of the EMI middleware and it is foreseen that the first concrete results will materialize after the first two major EMI releases. Therefore most of the industrial collaboration tasks are planned to take place during the second and especially third year of the project. A number of important collaborations have already been established to start the business model analysis and understand the issues involved with technology transfer, including licensing and copyright issues. In particular EMI is working together with Google to understand how the emerging dynamic service provision models based on open and commercial clouds can be used together with the traditional grid services. An MoU has been signed between EMI and Google to work on both the activities mentioned above. The first results of this collaboration are expected to be available at the end of the first year.
6 CONCLUSIONS

This document describes the initial definition of the EMI Strategic and Technical Roadmap. The roadmap includes a number of important milestones to be achieved during the project duration in order to fulfil the vision commonly defined by the EMI project partners and other relevant projects like EGI and PRACE.

The collaborations described in the Part 1 of the document will be established during the first two years of the EMI project, starting with EGI and the other DCI projects and progressively expanding internationally. A necessary condition for the definition of common objectives and the establishment of formal MoUs is the availability of the first release of the EMI distribution in April 2011, which many potential European and international partners see as the necessary first benchmark to assess EMI capability to implement its stated program of work.

It is understood that technology and user requirements keep changing at a very fast pace. EMI therefore needs to complement the initial definition of the roadmap with a careful analysis of new requirements and close collaborations with users and infrastructure providers. The roadmap will be re-assessed periodically and at least once per year to make sure that the milestones are achieved, but also that they are still relevant. Modifications and corrections will be proposed and the plan adapted to keep the EMI effort focused on the primary EMI objectives, which is providing users and infrastructures with efficient working solutions to their evolving scientific research needs.
PART 2

DISTRIBUTED COMPUTING INFRASTRUCTURE (DCI) COLLABORATIVE ROADMAP

Contributors on behalf of their projects: Steven Newhouse (EGI-InSPIRE), Alberto Di Meglio (EMI), Alexander Papaspyrou (IGE), David O'Callaghan (StratusLab), Jozsef Kovacs (EDGI), Andrea Manieri (VENUS-C)

Note: This material has been provided by the ‘Distributed Computing Infrastructure (DCI)’ projects funded in the November 2009 call to form a common text agreed by all the projects. This text is to be used unaltered by these projects in their respective deliverables. It is expected that the text will be placed into project specific document templates and into each project’s own review process to gain the endorsement of each project. If used as part of a document addressing other aspects it must either be referenced, or included in its entirety unaltered as an appendix in the document. This document has been produced with the co-funding of the DG Information Society & Media, European Commission. The content of this publication is the sole responsibility of the DCI project members and cannot be considered to reflect the views of the European Commission.
1 EXECUTIVE SUMMARY

As a result of an open call that closed in November 2009, six projects are being granted with nearly €50M by the European Commission’s 7th Framework Programme in the area of Distributed Computing Infrastructures. Together these projects will provide a pan-European production infrastructure built from federated distributed resources, ensure the continued support, maintenance and development of the middleware (gLite, ARC, UNICORE and Globus) that are in common use in Europe, explore how grid sites and different applications can be hosted sustainably in commercial, public, publicly procured and private ‘cloud computing’ environments, and provide desktop resources to the European research community.

These projects are the result of over a decade of community building that has taken place in the area of European Distributed Computing Infrastructures – both in their operational provision to a multi-disciplinary user community and the research and associated software development to build such infrastructures. Together, these infrastructures face the challenge of evolving their services to the changing needs of their data-intensive user communities, and providing a sustainable service that will support their users today, tomorrow and the years to come. A vision is presented of moving from the current production infrastructure in Europe to one based upon federated virtualised resources. It is expected that this change will increase the flexibility of resource providers to meet the changing needs of the user communities they serve by adopting best practices from other sectors.

In this report, the individual interactions between the six projects are recorded. Many of the projects expect to define these interactions through Memorandum of Understanding, and where there is an operational relationship between the projects through a Service Level Agreement. Not all of the projects have identified concrete interactions at this point in time with any other, though these may develop during the course of the individual projects.
2 INTRODUCTION

Call 7 (FP7-Infrastructures-2010-2) under the e-Infrastructures topic of the FP7 "Capacities" Specific Programme which closed in November 2009 called for proposals under the topic of ‘Distributed Computing Infrastructures’. As a result of this call, six projects were funded, with an expected total EC contribution of nearly €50M, that together are referred to in this and related documents as the ‘Distributed Computing Infrastructure Projects’ or ‘DCI Projects’. These projects are described in detail in Annex A and are summarised below:

- EGI-InSPIRE: Federation of national and domain specific resource providers into a European Grid Infrastructure for multi-disciplinary use.
- European Middleware Initiative (EMI): Continued support, development and harmonisation of the European middleware stacks from gLite, ARC, UNICORE and dCache.
- Initiative for Globus in Europe (IGE): Dedicated support for the European Globus community.
- European Desktop Grid Initiative (EDGI): To deploy desktop grids and cloud computing services for European user communities.
- StratusLab: Exploration of running production grid services in a cloud environment and providing cloud resources to research user communities.
- VENUS-C: Will explore and demonstrate the applicability of private and public cloud computing environments to different scientific applications to speeding up e-Science built on the sustainable public procurement of computing and storage resources on the cloud market.

Together the projects fund different activities in the area of distributed computing infrastructures ranging from the provision of production environments, the development, maintenance and support of the middleware used in Europe, and the exploration of the provision and use of virtualised computing resources. Due to the competitive nature of the funding model, many of these projects were developed in isolation in order to maintain confidentiality during the proposal phase. As a result, one of the goals required by the European Commission is that each project has to establish how they plan to collaborate with each other (if at all) and what the results of those collaborations may be within the scope of the project, and the impact that there may be long-term within the community.

This document shows how the provision of e-Infrastructures in Europe could evolve over the next 5 years and the contributions that each project may make towards this future by working with each other. As background, an overview of each of the six projects is provided in an Appendix.

It provides a record to the European DCI community as the potential results of the collaboration between the six distinct activities and the opportunities for collaboration it not only opens up between these projects but the wider community. It is essentially a technical document – describing the relationship between the projects and the technologies they will produce – and will it is expected become the basis for dissemination material to other interested stakeholders.
3 A VISION FOR EUROPEAN DCIS

3.1 WHAT ARE DCIS?

Historically, a single data processing or generating resource (storage, computers, instruments, etc.) has been under the exclusive control of the administrative domain that owns it. However, some scientific, academic and research organisations, which already own these data related resources, increasingly need to securely share these resources with others. In order to federate their local resources into a production infrastructure, these organisations have had to establish mutual trust, adopted compatible middleware stacks and procedures integrated through operations teams to bring their resources together into a distributed computing infrastructure (DCI).

The recurring feature of the various DCIs that offer production resources (e.g. EGI, DEISA, PRACE, etc.) is that each one integrates multiple locally managed administrative domains into a usable environment. The middleware deployed by each DCI provides its users, according to their DCI credentials, consistent access rights to all resources managed by that DCI.

3.2 A DECADE OF COMMUNITY BUILDING

The last decade has seen an unprecedented period of experimentation and prototyping in the collaborative use of distributed computing infrastructures. The EC funded European Data Grid (EDG) and Enabling Grids for E-sciencE (EGEE) projects have built a collaborative infrastructure of primarily High Throughput Computing (HTC) resources to support intensive data analysis. The DEISA and other projects have focused on integrating an infrastructure of High Performance Computing (HPC) resources to support large-scale computing simulations. Together, with the provision of an integrated network of National Research and Educational Network (NREN) providers supported through the GEANT series of projects, these activities have been developing the core of a European e-Infrastructure service.

These EC funded projects have also provided a structuring effect in the geographical region around Europe. Infrastructure projects such as BalticGrid and SEEGrid linking the Nordic and Baltic states and South East Europe. As a result of this activity many of these countries are now part of the EGI-InSPIRE project. This structuring relationship between Europe and other regions around the world continues in the networking, computing and application space through several related projects.

The contribution from the European Commission to this activity has been a small but enabling contribution to the investments made by the national funding agencies. The EC investment has contributed towards the staff needed to bring these compute, storage and networking into a European infrastructure. The hardware and operating costs for these activities, in addition to funding the research undertaken on the e-Infrastructure, has all been funded outside the FP7 programme. For instance, in the 4 years of the EGI-InSPIRE project, the EC investment of €25M to the provision of a European Grid Infrastructure is a small proportion of the estimated €330M invested by the countries involved in the project providing the European e-Infrastructure.

The provision of the European e-Infrastructure has been driven by the needs of the user communities that have needed access to large scale data analysis infrastructure to support their research needs as part of their pan-European research collaborations. Over the last decade, the European e-Infrastructure has benefited greatly from the growing maturity of the available open-source software solutions and where necessary have through middleware consortia such as gLite, UNICORE, ARC and Globus, and specialised technology providers like dCache, developed new, or extended existing solutions, in order to meet the needs of its user communities. These early adopting user communities have helped drive the development of the e-Infrastructure we have available today, which provides a production quality
federated resources, integrated through the middleware specifically developed to meet the demanding use cases coming from within the user and operations community.

3.3 CURRENT CHALLENGES

Even with the globally recognised achievements of the e-Infrastructure activities in Europe - delivering a production quality environment that supports a multi-disciplinary user community - broader adoption of e-infrastructures across the whole research computing community remains elusive. The reasons that other communities have not adopted the current e-Infrastructure offerings may range from:

- The data analysis challenges being faced by other communities have to date been within the scope of their current resources
- The usability and integration of non-local resources when compared to their desktop for solving problems is too high a barrier to overcome
- The service offering developed for the current user communities do not match the needs of other communities
- The future sustainability and governance of the e-Infrastructure to those communities that have not been actively involved in its development is not clear or assured
- The true cost of delivering a world-class data-intensive analysis infrastructure, regardless of the resources used to deliver it, needs to be exposed to the resource providers, the consuming end-user community and policy makers.

Recent activities within Europe are addressing these five issues.

The next generation of pan-European research infrastructures (the projects that are part of the European Strategic Forum on Research Infrastructures – http://ec.europa.eu/research/esfri) presents an opportunity for European e-Infrastructure providers to support a new wave of data-intensive research activities that will be highly dependent on a distributed computing and storage models. For these new communities, establishing and maintaining their own independent e-infrastructure is a diversion from their primary mission of doing science. Having access to a reliable European e-Infrastructure, available as a service, becomes an attractive option. It is also essential that any e-Infrastructure that they use in Europe be integrated with the e-Infrastructure used by their non-European research collaborations.

After a decade of investment in European e-Infrastructure, the production quality service offering now provided to the European Research Area are coalescing into two main areas:

- High Performance or Capability Computing provided currently by the DEISA and in the future the PRACE (Partnership for Advanced Computing in Europe) projects integrate high-end resources (generally of 10,000-100,000 cores) across Europe. Generally, these resources are used for closely coupled parallel applications for the few researchers with problems and applications able to benefit from them.

- High Throughput or Capacity Computing integrated into a European Grid Infrastructure (EGI) supported through projects such as EGI-InSPIRE and EDGI. These resources may include loosely or tightly coupled clustered, volunteer desktop or virtualised computing clusters contributed into a European infrastructure through national groupings of resource providers. Generally, these resources support the ‘bags of tasks’ applications where each task involves the execution of a program which needs minimal porting to run in such an environment.
The discussion for the remainder of this section will concentrate on the development of ‘high throughput’ computing resources both for single processor and parallel applications which are expected to provide the majority of resources and support the main stream of application communities in the years to come as new virtualised computational resources (currently available commercially on demand as cloud computing resources) are evaluated for integration into the publicly funded production infrastructure in Europe.

3.4 EGI AND THE DCI EVOLUTION

The goal of EGI is to provide a secure integrated federated production infrastructure constructed from national and domain specific resource providers, that is open to all users with potentially different computing models, needing access to different types of distributed resources (high-throughput, high-performance, desktop, virtualised, etc.), that are linked to physically remote data stores. Some of the high-performance computing resources may include some of those currently classed as DEISA resources.

Such an environment - a secure integrated sustainable production infrastructure - imposes constraints on those that produce software technology for deployment within it, and those that provide the resources to the infrastructure.

The EGI model is based around the contribution of resources from within different administrative domains where remote access is given to defined virtual organisations (groups of individuals coming together for a common goal) which will include users from different organisations. At the core of this model must be common mechanisms for establishing and entity’s identity (authentication) and the ability to control access to particular resources (authorisation). These mechanisms must be embedded into the access mechanisms for all resources to ensure a consistent reliable predictable security model.

Integration is necessary so that end-users are presented with consistent reliable secure interfaces to the same class of resource regardless of the resource provider and the implementation used to expose the resource to other users. To give resource providers the ability to deploy different software implementations to provide the same functionality it is necessary that the implementations demonstrate interoperability. The easiest route to achieve this is through the adoption of standards, and the verification of these interfaces through appropriate conformance tests. The ability of the interface deployed on a site to be available and to behave as expected is an aspect that is monitored remotely. High availability and the planned management of outages (reliability) is a vital aspect in defining a production (as opposed to a research) infrastructure.

In addition to monitoring the availability and reliability, an additional characteristic of a production infrastructure is the ability to account for its usage. This is important in order to understand current usage (of sites by virtual organisations) and to plan for changes that may occur in the future. Resources in a production infrastructure must therefore provide accounting information that allows usage by resource, resource type and virtual organisation to be tracked.

Against these technological constraints is a need for a sustainable operational model to be developed. The user communities planning to adopt the production infrastructure to support their research activities are doing so as part of a research programme that may persist for 10 or 20 years. Therefore the sustainability of the resources, the way they are funded, organised and operated, has to persist outside of any particular project. The European Grid Infrastructure (EGI) is now coordinated on behalf of the community by a dedicated organisation (EGI.eu) funded and responsible to the community it serves. It provides the centralised coordination necessary to bring together individual resources providers, either public or private, to deliver a secure integrated infrastructure, and as a means to gather and prioritise requirements from the resource providers and user communities as to how the infrastructure should develop.
3.5 THE DCI COMMON VISION

The pressures (staffing costs, green energy, economies of scale etc.) that produced the consolidation of data centres and wide-scale adoption of virtualisation in the commercial sector are beginning to be felt in the academic and research space. Many campuses are encouraging the move of departmental or group level computing resources into central locations where they can be managed and supported by dedicated staff. This trend will inevitably continue over the next decade, forcing a greater integration between the client environment available at the researchers fingertips and the remote resources that they have access to ‘somewhere’ over the Internet. The ‘where’ of these resources will become increasingly less important to some communities, but of critical importance to those where their data is governed by legislation (e.g. medical, personal, financial, etc.). A researcher will have access to a pool of resources that are available to them through their roles within physical organisations (e.g. their employer), their funders (e.g. national resources), through their collaborations (e.g. international virtual organisations) or acquired commercially. Much more important will be the ‘how’ of configuring and exploiting these resources effectively for their own needs or those of their collaborators.

This ability to provision resources ‘on-demand’ to meet the needs of particular research collaboration provides significant challenges to resource providers in the research space. In the commercial world ‘cloud computing’ has provided a ‘pay per use’ business model that has shown the use of virtualisation to deliver ‘Infrastructure as a Service’, hosted environments to provide a ‘Platform as a Service’ and hosted applications to access ‘Software as a Service’. Cloud providers offering Infrastructure as a Service can be integrated seamlessly alongside the academic resource providers offering a virtualised compute resource – but currently without the direct integration with the GEANT network.

Figure 5 - Virtualised Federated Resources
In the research world the trend to consolidate data centres will also be continued to be balanced with the need to collaborate to share resources. This will lead to federated grids of virtualised resources (See Figure 1), in many ways similar in architecture to today’s federated grids of computing resources, but providing truly generic infrastructure that can be accessed by any authorised research collaboration, as an alternative or alongside commercially provided resources. This virtualised infrastructure will be used to instantiate a platform to support particular research collaborations. These platforms (which will be comparable to the current gLite, Globus, ARC or UNICORE environments) may be deployed directly by the research collaboration using pre-defined images, by using bespoke images created from within the collaboration, or provided as a service by third-parties within the ecosystem. To many of the end-users within the research community who take no interest in the details as to how their infrastructure is provided, the result will appear as just a set of services available ‘out there’ for them to use.

3.6 IMPLEMENTING THE VISION

A grid of virtualised resources, with federation taking place within a region, national borders, or across the European Research Area has many potential benefits, opportunities and challenges for end-users in the research community, service providers within the research community, commercial organisations wishing to engage and provide services and resources to the research community, and other organisations for establishing the policy environment for such a federated infrastructure to operate within.

The following sections discuss some of the challenges and potential benefits of the proposed model for some of the participants within this vision for DCI provision within Europe.

3.6.1 For Infrastructure Providers

This roadmap for European DCI provision provides many benefits. The alignment of infrastructure provision in the research e-Infrastructure community with models used in the commercial world - provision of end-user environments through virtualisation - allows tools and techniques used in industry to be adopted in academia. This approach has already demonstrated increased server utilisation, better energy utilisation and greater flexibility in the commercial world. Tools developed to meet the management challenges used in these commercial environments may also prove effective in the research environment.

This virtualised environment will allow resource providers to deploy virtual images on demand to meet the needs of different user communities. The flexibility provided through a trusted repository of virtual images would allow resource providers to support a greater number of different environments and therefore a greater number of different user communities. This provisioning activity may be undertaken directly by the local resource provider or by authorised third-parties, e.g. from other resource centres, by representatives of the user communities or by other authorised entities. Such a model requires a trust model between the local resource provider and the generator of the virtual image. Policy and technical discussions around this area are ongoing and conceptually such a trust model is similar to that currently used for the pilot job frameworks used within the High Energy Physics community where the resource provider delegates the actual payload executed in their machine to a trusted third party. Implementing such a model requires work within the community to manage the distribution of virtual machine images, mechanisms for image signing and site policies for accepting images based on signed images to create trusted image repositories.

A virtualised infrastructure that allows environments to be deployed on demand by authorised groups allows a different security model to be used for the provisioning activities than is used within the virtual machine image. Clearly, there is a need for the infrastructure provider to be reassured as to the activity that will take place within the instantiated virtual machine, depending on what that activity might be. The level of authorisation and logging that may be needed within the virtual machine might
be conditional on the end-user control of what takes place within the virtual machine. A similar conditional policy on the levels of authorisation is in place for portal access to the e-infrastructure.

### 3.6.2 For the Software Provider

Large-scale adoption of virtualisation by the infrastructure, and the effective management of the software deployed within the virtual images, imposes operational requirements on the software services. Once the virtual machine is running consistent service management and monitoring interfaces are needed to configure the services within its instantiated environment (e.g. What services should be run? What certificate should be used? Who is allowed to access these services? etc.) and to monitor their operation and health. Providing consistent standardised interfaces enables third-party management tools and protocols to be used to support manual intervention by the operations staff.

The loosely-coupled dynamic nature of this infrastructure needs a flexible system for linking the virtual machine hosting environments, the transient virtual machine images that run on the hosting environments and the services within the virtual machine image itself. Modern messaging systems have been designed for use in just such a distributed environment through the ability to have different messaging queues and provide persistent message delivery. A messaging infrastructure will underpin the future DCI and should be used as the basis for messaging and management by the deployed services.

The messaging infrastructure provides a basis for higher level applications to build upon. This includes existing operational functions such as accounting and service monitoring, and provides a basis for research into new operational tools such as autonomic management of the infrastructure. As the scale and complexity of the infrastructure increases, autonomic management functions become essential - to recognise when virtual machine instances or services have stopped working and to restart or redeploy these instances to ensure the required services remain available to the user communities.

### 3.6.3 For the end-user

For end-users in the research community, easy usage of e-Infrastructures is essential – regardless as to who operates it or the technology used to deliver these services. A federated virtualised infrastructure presents many additional benefits. It provides a means for the user communities to deploy within the infrastructure the services that they wish to use when they wish to use them. These services will need to be encapsulated into a virtual machine image and be able to meet the policy requirements imposed by the infrastructure relating to security, configuration, management, monitoring, etc.

Provisioning of this infrastructure for end-user communities may come from many sources. For communities that have the required technical knowledge they may generate their own customised virtual machine images to the appropriate specifications and deploy these to the virtualised resources that they need to use. Other communities may work with experts outside the community to have an environment created, deployed, managed and monitored on their behalf. The resource providers may provide a basic generic environment over some resources to provide a resource for communities that do not need a customised environment. The execution environments required by an end-user could also be made available through commercial cloud infrastructures alongside other resources that are able to provide dynamic scale-out capabilities, which do not require long-term resource and organisational commitments. This model provides much greater flexibility to the end-user community as to the environment that is available to them - if they need to exploit such flexibility.

The use of messaging as a fundamental part of the infrastructure provides flexibility in how the end-user interacts with the distributed resources. For instance, it provides the ability for the user to easily subscribe to events that they are interested in - when an application starts running, when it stops, or if it fails to complete. Results can be sent back to the user through the messaging system. As the messaging system is capable of asynchronous delivery, it allows results to be stored and then
delivered when the user is ready - for instance when they reconnect their laptop to the network in the morning.
4 INDIVIDUAL INTERACTIONS

To support the development of e-infrastructure provision in Europe and the vision outlined in the earlier sections the DCI projects will be working on various collaborative aspects. These are summarised in the table below and detailed in the remainder of this section.

4.1 EGI-INSPIRE

The EGI-InSPIRE project’s main focus is to deliver a production infrastructure for the European Research Area. In order to deliver this, it needs to deploy a software environment that brings together software components provided from both within and from outside the DCI community.

Two projects are seen to initially provide these software components:

- European Middleware Initiative (EMI)
- Initiative for Globus in Europe (IGE)

A Memorandum of Understanding (MoU) will be established with each project to describe common plans around dissemination, representation to ensure the exchange of requirements and the development roadmaps. Specific Service Level Agreements (SLAs) will be defined to govern the expected operational interactions on the provision of third line support and security incident handling.

It is envisaged that future DCIs will make extensive use of virtualisation technology. An MoU will be established with StratusLab to ensure effective joint dissemination and events, where applicable, and for the operational staff within EGI to ensure that the software environment being released from StratusLab will be the needs of the production infrastructure. This may include requirements on reliability, scalability, monitoring and accounting.

The main output from VENUS-C will be a series of user scenarios showing how the cloud computing model can benefit different scientific communities. VENUS-C will expand the supported communities by mean of an open call for up to twenty short experiments to exploit the VENUS-C Cloud Platform through the cloud resources provided within the project. EGI-InSPIRE would like to provide input to VENUS-C on the criteria for the experiments that are selected from the open call, to ensure they are of relevance to the EGI user community.

It is expected that the EDGI project will build a desktop resource across Europe. EGI-InSPIRE would like to ensure that this computing resource can be integrated alongside the resource types offered within the production infrastructure. For this end, EGI-InSPIRE will collaborate with EDGI through an MoU that will establish the monitoring, accounting and functional integration of desktop resources into EGI. As a result of this the reliability and use of this resource can be established alongside the other provided resource types and these desktop resources can then be offered up to the EGI user community alongside the others.

4.2 EMI

As one of the major providers of middleware services for the Distributed Computing Infrastructures, the EMI project will establish interactions with both infrastructure providers deploying the services and other middleware and application developers complementing or extending the EMI services. In particular, within the group of DCI projects described in this document the following interactions have been identified after the first four months of operations of EMI in numerous discussions with the relevant projects.
Table 2 DCI Interactions summary

<table>
<thead>
<tr>
<th>Providing Project</th>
<th>Consuming Project</th>
<th>EDGI</th>
<th>EGI-InSPIRE</th>
<th>EMI</th>
<th>IGE</th>
<th>StratusLab</th>
<th>VENUS-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDGI</td>
<td>Dissemination and integration of desktop resources.</td>
<td>Technology bridging through interoperability and standardisation.</td>
<td>Interest from UK NGI, EDGI-SW, GRAM gateway development, requirements.</td>
<td>Requirements for integrating virtualisation into the operational infrastructure.</td>
<td>Adapting desktop grids to be run over cloud resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGI-InSPIRE</td>
<td>Dissemination.</td>
<td>Deployment, requirements, dissemination, feedback &amp; usage.</td>
<td>Deployment, requirements, dissemination, feedback &amp; usage.</td>
<td>Requirements for Globus, standardisation and interoperability.</td>
<td>Promoting Open Call toward EGI user communities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMI</td>
<td>Technology previews.</td>
<td>SLA defined middleware maintenance and support. Dissemination and training.</td>
<td>Requirements for Globus, standardisation and interoperability.</td>
<td>Middleware able to run on OpenNebula.</td>
<td>Middleware able to run in VENUS-C (under defined constraints).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGE</td>
<td>Globus support and GRAM gateway development.</td>
<td>SLA defined middleware maintenance and support. Dissemination and training.</td>
<td>Standardisation and interoperability. Support for Globus components.</td>
<td>Support of StratusLab Cloud Platform, support globus.eu, investigate dynamic Grid deployment.</td>
<td>Adapting globus-enabled application to run over cloud resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StratusLab</td>
<td>Virtual appliances for Grid services</td>
<td>Requirements dynamic deployment of virtualised grid sites.</td>
<td>Requirements for Virtual appliances for Grid services, access to virtual testbeds</td>
<td>Hosting of globus.eu (using Amazon API) &amp; requirements for dynamic Grid deployment</td>
<td>Adopting Stratuslab toolkit as IaaS middleware in one or more VENUS-C sites.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VENUS-C</td>
<td>Supporting desktop grids on top of cloud resources.</td>
<td>Best practices of scientific communities using clouds. Opportunity for EGI communities to experiment with the VENUS-C platform through the Open Call</td>
<td>Requirements for Virtual appliances for Grid services, access to virtual testbeds, Sharing information on accounting data formats and approaches</td>
<td>Supporting Globus users in experimenting on cloud resources.</td>
<td>Providing feedback on usage requirements and user experience on the IaaS approach.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2.1 EGI-InSPIRE

The European Grid Infrastructure, supported by the EGI-InSPIRE project, is the main user of the EMI services. EGI represents therefore the major source of requirements and the primary target for the delivery of software and support services from EMI. EMI and EGI have discussed in several occasions the points of interactions and a common vision and common plans for maintaining and evolving the European research infrastructures. The relationships between EMI and EGI will be formalised with the establishment of a Memorandum of Understanding and the negotiation and signature of a commercially-oriented Service Level Agreement. In particular the SLA will describe in details how the two parties commit to provide and access the EMI services and what levels of service quality are expected. It is EMI intention to establish together with EGI a prototype of possible future professional service provision relationships that could extend to commercial providers.

In summary three major collaboration points have been defined:

**Requirements**: the collection of requirements is an essential part of the middleware development lifecycle. EMI must be sure that what is provided by its Product Teams is relevant, usable and able to support the EGI roadmap vision for the research infrastructures. EMI will therefore actively take part in the definition of the EGI UMD Roadmap by means of mechanisms provided by EGI, like the Technology Collaboration Board (TCB). EMI will provide clear deadlines for the releases of its software services with particular attention to the existing and new functional requirements discussed with EGI and its user communities. In addition, EMI will work with EGI on the definition and implementation of the Infrastructure Roadmap and the integration of existing middleware services with emerging computing and data storage technologies.

**Maintenance and Support**: Although EMI has to evolve the middleware services towards the implementation of the DCI vision, the continuous operational efficiency of the infrastructures has to be guaranteed. EMI together with EGI will continuously monitor and assess the quality of the software developed by EMI and deployed by EGI in order to react to any incident or user request in the most professional manner. Clear criteria for transitioning the services from EMI to the EGI roll-out service will be defined and periodically revised. Clear support policies, service lifetime policies and migration paths from old to new services will be defined by EMI and EGI and formalised in commercial quality Service Level Agreements, which will be periodically revised and improved. EMI and EGI together will also explore possible alternative model for supporting the middleware involving commercial partners whenever feasible and desirable.

**Dissemination and exploitation**: EGI and EMI together represent a large part of the European and international infrastructures in support of scientific research communities. An efficient and timely dissemination of information and the expansion of the user base are keys to the correct exploitation of the infrastructures. EGI and EMI have therefore engaged in establishing common dissemination strategies to provide coherent and complete perspectives on the various components of the infrastructures and their applications. A first important result of this engagement will be the joint organization of a major international event during Spring 2011 to bring together existing and new scientific user communities, presents achievements and results and collect trends and ideas.

4.2.2 IGE

The Initiative for Globus in Europe is a provider of middleware aiming at creating an official link between European users of Globus and the US developers maintaining it. A number of Globus components are currently used within the EMI services to provide specific functionality. In addition, a number of services developed by EMI and Globus are providing similar, but not always interoperable functionality. EMI and IGE have therefore identified two major areas of collaboration, which will be formalised with an appropriate MoU:
Standardisation and Interoperability: EMI and IGE will work together and as part of other standardization and interoperability bodies to defined and evolve standards for the distributed computing middleware, especially in the areas of Compute and Data Management.

Support and maintenance: EMI needs to rely on continuous support from the Globus developers and maintainers in case of software issues. EMI will also provide requirements for Globus to IGE as needed and will monitor together with IGE the implementation of those requirements in the Globus releases. In exchange EMI will gradually move from its current usage of Globus, distributed as part of the EMI middleware services, to a more standard use of official Globus packages maintained by IGE and distributed as part of the major Operating Systems distributions, like Fedora or Ubuntu.

4.2.3 EDGI
EMI and the European Desktop Grid Initiative are software providers for EGI with strongly complementary roles, since they maintain and promote different types of distributed computing middleware for different sets of use cases. There are a number of interesting common points that link EMI and EDGI at the boundary where the two technologies meet. Essentially part of the EDGI services makes use of middleware services provided by EMI to bridge standard grids and desktop grids. The work of bridging the two types of grid was already started in previous project, but EMI and EDGI are now planning to work together in completing such bridges and providing access to resources not only via gLite, but also via ARC and UNICORE and future standard-based resource management clients. EDGI will work with EMI as part of the ‘Works with EMI’ technical collaboration program that allow technology providers and consumers to have direct access to technical previews and dedicated support for complementing and extending the EMI services.

4.2.4 VENUS-C
VENUS-C is providing both platform APIs and resources to enable scientific users to access commercial cloud providers or (public or private) data centers. As part of the overall DCI vision for how the research infrastructures will be shape in the coming years, it is clearly acknowledged that cloud or similar dynamic service provision models will be more and more used. EMI is therefore fully committed to understand how the existing distributed services can be improved and evolved to exploit such service provision models while retaining their existing flexibility and security. EMI and VENUS-C are discussing on the possible integration paths across grid and cloud platforms. EMI will put effort in introducing any modifications in it services to make them fully compatible with the VENUS-C infrastructure. At the same time VENUS-C will provide EMI with access to technology and resources to validate and test the EMI services on virtualized environments. Common work on security and accounting formats enabling interoperability will also be considered.

4.2.5 StratusLab
StratusLab is providing software to setup distributed computing infrastructure based on the emerging cloud technology. As part of the overall DCI vision for how the research infrastructures will be shape in the coming years, it is clearly acknowledged that cloud or similar dynamic service provision models will be more and more used. EMI is therefore fully committed to understand how the existing distributed services can be improved and evolved to exploit such service provision models while retaining their existing flexibility and security. EMI and StratusLab are discussing on how existing EMI grid services can run on virtualized environments based on OpenNebula and which modifications are needed in the services configuration capabilities to make them able to be instantiated as on-demand services or pre-configured appliances. StratusLab will provide EMI with requirements and with access to testbeds, while EMI will incorporate and support in its releases the functionality needed to exploit virtual environments based on StratusLab technology.
4.3 IGE

The IGE project strives to integrate as tightly as possible with the other DCI projects in order to deliver a convincing user experience to the scientists within the European Research Area. To this end, interactions are to be established as follows:

4.3.1 EGI-InSPIRE

The main goal is to ensure collaboration for the integration activities with respect to Globus-contributed infrastructure, wherever appropriate. This includes acting as a software provider towards EGI by setting up reasonable SLAs, delivering Globus and Globus-related components to the UMD, and contributing training and support where necessary. Moreover, IGE will ensure the appropriate representation of European Globus-based research communities within the Virtual Research Environments.

As a first concrete action, a Memorandum of Understanding (MoU) will be established that describes common plans around dissemination, representation, and exchange of requirements and development roadmaps.

4.3.2 EMI

IGE will strive to become the main provider of Globus components within the EMI software stack. To this end, a close collaboration with EMI will be setup in order to ensure continuous support for the US-based Globus package distribution, and to collect additional requirements from EMI, such as more standard installation and deployment procedures (through major Operating System distributions such as Fedora and Ubuntu). Moreover, both projects will collaborate in the area of standards and interoperability in the areas of Compute and Data Management for DCI middleware.

As a first concrete action, the modus operandi of this interaction will be detailed in a MoU between EMI and IGE.

4.3.3 StratusLab

One major goal will be to work with StratusLab on hosting the anticipated globus.eu branch of the newly developed Globus.org SaaS platform on StratusLab infrastructure. In this context, the dynamic deployment within a Grid environment (i.e. submitting VMs instead of traditional jobs through a Compute Service interface such as GRAM). Moreover IGE will interact with StratusLab to ensure that (a) the StratusLab Cloud Platform by itself and (b) the new delivery paradigms are supported within Globus; in particular, it is to be ensured that Globus is compatible with the IaaS interfaces and the creation of VMs.

To this end, IGE will frequently test the StratusLab innovations with Globus and regularly provide feedback throughout the whole development.

4.3.4 EDGI

Recently, NGI-UK has formulated the requirement to utilise Desktop Grid resources as part of the national infrastructure. Since Globus is part of the middleware stack here as well, IGE will collaborate with EDGI to develop, contribute, and maintain a Desktop Grid Bridging Service for Globus to cater the need for EDGI integration.

4.3.5 VENUS-C

IGE offers to support VENUS-C in understanding the special requirements of traditional Grid users for enabling their applications to use Cloud infrastructures, how such use cases can to be deployed, and by providing access to Globus, Grid application, and test resources. IGE will also provide technical assistance and expertise to VENUS-C where required on pertinent aspects of platform interoperability with Globus, e.g., concerning Security and Authentication.
4.4 EDGI

As a collaboration activity between EGI.eu and EDGI, the EGI.eu dissemination and training channels will be used to reach the existing EGI user communities. In order to get the highest possible impact EDGI will organize dissemination events in the framework of events organized by EGI.eu. International Desktop Grid Federation and EDGI will work in strong collaboration with EGI, EMI, NorduGrid, UNICORE Forum and interested NGIs in order to reach the widest possible user and resource provider communities. Most of the dissemination work of EDGI will be done in the framework of the European Chapter of the International Desktop Grid Federation.

The Desktop Grid Federation is set-up to be long-lived, i.e. after the EDGI project has finished. We try to align it as much as possible with existing e-Infrastructure organisations, such as EGI, so it could also be possible that (part of) the Federation could become a user group in these e-Infrastructure organisations. The International Desktop Grid Federation will organise the grid operators and application developers in the European Union. It will strongly collaborate with the International Desktop Grid Federation run by DEGISCO in order to organise the grid operators and application developers outside the European Union, especially the ones in the ICPC countries.

Standardization activities will be carried out through several channels. The EDGI Bridge will use the HPC profile job submission mechanism (an OGF standard) in order to guarantee the interoperability with every Service Grids that follow this standard. In particular, the UMD developed in EMI currently follows this standard that makes sure that the EDGI Bridge will be compatible with the middleware supported by EGI.eu. In order to maintain this compatibility for the whole duration of the project and beyond, EDGI will strongly collaborate with EMI.

EMI will concentrate on the major Service Grid middleware systems and will further develop ARC, gLite and UNICORE towards making them interoperable and based on them will create a unified middleware distribution, but will not cover any Desktop Grid extension of these middleware systems. EDGI will cover this important area in the e-science infrastructure eco-system. The objectives of EDGI and EMI are complementary (both want to further develop middleware) but technologically orthogonal (the middleware to develop are different). EDGI will carefully follow any improvements and further developments of ARC, gLite and UNICORE created by EMI in order to make sure that the Service Grids → Desktop Grids bridge middleware developed by EDGI will be compatible with any new versions of the ARC, gLite and UNICORE middleware.

EDGI will strongly collaborate with DEGISCO that is a support action project to disseminate the results of the EDGeS project outside the EU countries. Since EDGI is also a follow-up project of EDGeS and aims at disseminating desktop grid related knowledge in EU countries there are many commonalities between the two projects.

EDGI is furthermore open for any DCI project to use Desktop Grid resources. One possibility is to provide a solution for Globus users to transparently and seamlessly utilise Desktop Grid resources through the EDGI Bridge. To support this idea, EDGI will investigate the possible alternatives together with the IGE DCI project. One potential user for the Globus → Desktop Grids Bridge is the UK NGS.

4.5 STRATUSLAB

StratusLab is open to collaboration with all DCI projects with an interest in using cloud resources.

4.5.1 EGI-InSPIRE

StratusLab will deploy grid sites over cloud infrastructures that will join the EGI infrastructure. The operation of virtualized sites may require adjustments in the way grid resources are certified, managed
and operated. A virtualized grid site will expose elasticity and volatility at a level not previously experienced in operational sites.

The two projects will have to collaborate closely in order to ensure that the operational models implemented by EGI will be cloud-friendly and flexible enough in order to take advantage of the merits brought by cloud computing. Additionally cloud software should be enhanced in order to enable dynamic provisioning and configuration of grid-resources permitting the provision of grid-sites-on-demand.

4.5.2 EMI

The European Middleware Initiative project is responsible for the support and development of the Unified Middleware Distribution (UMD) a term used to refer to an integrated distribution of the most popular European grid middleware, namely gLite, UNICORE and ARC. Apart from the existing high-level collaboration among DCI projects, there is also room for collaboration on a technical level. StratusLab will be using UMD to deploy grid services on top of cloud infrastructures. In many cases technical restrictions may impede the efficient installation and operation of these grid services. A channel of interaction among the two projects would be important in order to convey problems and requirements. The final goal of the above collaboration will be at the end of the projects to have UMD and the StratusLab software distributions to be fully compatible.

4.5.3 IGE

IGE will develop grid services and tools that can take advantage of the Cloud. StratusLab could provide to IGE requirements and an architecture for Grid sites taking advantage of Cloud concepts and technologies, and later cloud-enabling Globus-based infrastructures using StratusLab architecture and tools. IGE could provide to StratusLab requirements from Globus-based infrastructures, and later Globus services (e.g. GRAM for job execution) and tools (e.g. GridWay Metascheduler) able to operate on a StratusLab-based infrastructure and to take advantage of Cloud concepts. StratusLab and IGE could work together (probably at a technical level) on cloud-enabling of Globus services and tools using the StratusLab distribution.

4.5.4 VENUS-C

StratusLab and VENUS-C share a common interest in cloud computing. VENUS-C will develop and deploy a Cloud Computing platform service for scientific communities in Europe. StratusLab can provide VENUS-C with the StratusLab toolkit, as a comprehensive, open-source private cloud distribution for the VENUS-C IaaS backend to form part of the VENUS-C infrastructure on one or more sites. StratusLab will be interested in the requirements of VENUS-C for the IaaS platforms used in their infrastructure. Also, some partners of VENUS-C plan to evaluate OpenNebula as an IaaS solution, their feedback would be valuable. VENUS-C partners may provide adaptations of OpenNebula components made for their deployments. (e.g. development or tuning of the storage/network/virtualization plugins) StratusLab and VENUS-C may work together on the definition of extensions of current IaaS API’s.

4.5.5 EDGI

StratusLab will produce and maintain a repository of virtual appliances for grid services. EDGI may use these virtual machine images and supply related requirements,

4.6 VENUS-C

In scientific computing, there are user communities which traditionally have no strong need to utilize grid and supercomputer facilities. Instead, they utilized local smaller HPC compute clusters to do their simulations. In recent years, these user communities are tackling more and more complex
computational problems. The VENUS-C project’s goal is to equip these users with a tool set to easily scale their existing scientific workloads into the publicly available cloud resources. The expectation is that cloud resources are instantly provisioned, and that the users can scale out their workloads quickly at predictable costs. In order to keep the barrier of entry as low as possible, the project aims to keep the software dependencies and requirements as minimalistic as possible. A scientist should be able to test a scientific executable or script on his laptop, and then easily scale out the job in a map/reduce fashion to his resources in a public cloud.

The project will enable seven existing e-Science applications as part of the original workplan, as well as an additional of up to twenty selected applications from an open call, to run in the cloud. Several VENUS-C project partners donate a significant amount of both IaaS- and PaaS-based cloud resources (compute, storage, transfer) to the project’s scenario partners and the open call participants.

4.6.1 EGI-InSPIRE
As part of its open call ready for communication in the latter part of 2010, the VENUS-C project will financially support up to 20 selected applications to run their applications ‘in the cloud’, using both the VENUS-C tool set and the allocated cloud computing and storage resources. In order to obtain feedback from a broad and heterogeneous set of scientific applications, the VENUS-C project invites and encourages the EGI-InSPIRE project and the EGI user communities to engage actively in the open call. Broad participation in the open call will enable the VENUS-C project to assess the applicability of cloud computing to scientific communities, and a diverse set of applications will enable the identification of specific communities which may particularly benefit from the adoption of cloud computing. From an econometrics point of view any legal, and socio-economic findings on the suitability of adopting the cloud model for the EGI service provisioning could be shared between the projects.

4.6.2 EMI
EMI will provide a reference implementation for grid middleware in Europe. The trends in virtualised infrastructure and cloud computing will allow EMI to experiment such technologies and to better understand the cloud model. One opportunity for EMI and VENUS-C is to assess how traditional grid middleware (as supported by EMI) relates to a computational model put forward by VENUS-C. One concrete technological area for collaboration between EMI and VENUS-C is security and accounting. In particular, it would be interesting for VENUS-C to learn from EMI about the EMI project’s consolidated vision for accounting scientific users across different middleware platforms.

On the other hand, VENUS-C can support EMI in adopting both the cloud model and understanding the cloud technology, by providing access to best practices and resources to experiment and test the EMI services, at a cost of developing any bridge or slight adaptation of their services.

4.6.3 IGE
VENUS-C can support IGE in allowing the project, as a facilitator to Globus users, to understand how to adopt the cloud model and understanding the cloud technology, by providing access to best practices and resources to experiment and test Globus. With budget and resources permitting, IGE may look at developing an interface of Globus and VENUS-C. In addition, IGE may, as in EMI, also share information and experiences on the aspect of Authentication and Accounting in order to speed-up the interoperability amongst the different platforms.

4.6.4 EDGI
VENUS-C and EDGI may explore in the coming months the opportunity and feasibility of running desktop grids ontop of Cloud services, providing feedback both on the desktop grid applications and on the Platform APIs to increase the adoption of the Cloud environments by the scientific communities.
4.6.5 StratusLab

VENUS-C aims to deploy scientific workloads both on IaaS and PaaS cloud offerings. The StratusLab project aims to develop an IaaS StratusLab Cloud Distribution. Depending on the schedule and availability of the StratusLab toolkit, it might be possible to conduct an experiment to utilize a StratusLab-based grid resource from within VENUS-C. Given that OpenNebula is a relevant cloud platform in both VENUS-C and in StratusLab, it is desirable to exchange user and development experiences between the two projects. Therefore, VENUS-C and StratusLab may jointly promote the evolution and the adoption of standardised API at Infrastructure level, sharing best practices and interfacing with addressed standard bodies.
5 CONCLUSIONS

Even with the committed investment from the European Commission through the FP5, 6 and 7 programmes and the member states of the European Union, a clear challenge remains for European e-Infrastructure providers. To ensure their longer-term sustainability they need to be seen as providing a reliable and efficient service to all user communities in Europe needing to use research computing and storage services. This is essential in order to be able to attract a broad base of European and national research funds. However, in order to attract these user communities a broader range of services need to be provided for these individual communities within the same (and probably reduced) operational costs.

New technology offers a route for resource providers in the research sector to deliver these services with greater reliability, scalability and efficiency. However, such a route is not without its challenges. Firstly, it requires fundamental changes in how services are provisioned in the research community that builds on the experiences gained in the commercial space by using virtualised infrastructure to produce in some sectors a so-called cloud business model. Secondly, it requires open-source software community to adapt their software to be managed, monitored and deployed within a federated virtualised environment by focusing on delivering services that are not available elsewhere. Thirdly, that the user communities find the services offered to them attractive and easy to use so that they can be incorporated into their data analysis workflows. Finally, it provides some challenges to the computer science community to provide solutions that enable the reliable and effective management of such a highly distributed infrastructure.

Together, the DCI projects are able to address some of these issues and move the community towards the presented vision of an integrated virtualised infrastructure for the Europe Research Area. EGI-InSPIRE will provide a route for the deployment across Europe of new technological innovations into production once they have shown sufficient robustness and value to the EGI community. EMI and IGE provide a source of innovation in the short-term, and it is expected this will be expanded over time to include the technology and procedures developed within the StratusLab project. VENUS-C will provide best practices and potential success stories to the EGI community on the applicability of “cloud computing” for scientific computing, while EDGI will provide desktop and cloud resources to various European research communities.

The vision presented in this document transitions the DCI community to providing an integrated infrastructure as a service and for EGI to help bring the technology innovations being developed within the DCI community and elsewhere through to use in the research and public sectors. Interoperability and integration between different e-Infrastructures and technologies is fundamental to the DCI projects and the work that will be undertaken between them. This activity will help contribute to the broader vision described in the recent Digital Agenda for Europe (DAE)\(^4\) communication that “Europe should also build its innovative advantage in key areas through reinforced e-Infrastructures”.

---
\(^4\) A Digital Agenda for Europe : Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - COM(2010) 245, 19.05.2010
6 ANNEX A: THE DISTRIBUTED COMPUTING INFRASTRUCTURE PROJECTS

Each of the six Distributed Computing Infrastructure (DCI) projects are summarised in the following table and described in more detail in the following section.

<table>
<thead>
<tr>
<th>Project</th>
<th>EDGI</th>
<th>EGI-InSPIRE</th>
<th>EMI</th>
<th>IGE</th>
<th>StratusLab</th>
<th>VENUS-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start Date</td>
<td>01/06/2010</td>
<td>01/05/2010</td>
<td>01/05/2010</td>
<td>01/10/2010</td>
<td>01/06/2010</td>
<td>01/06/2010</td>
</tr>
<tr>
<td>Duration (months)</td>
<td>24</td>
<td>48</td>
<td>36</td>
<td>30</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Total Budget (€)</td>
<td>2,436,000</td>
<td>72,000,000</td>
<td>24,000,000</td>
<td>3,693,000</td>
<td>3,137,221</td>
<td>8,803,046</td>
</tr>
<tr>
<td>Funding from the EC (€)</td>
<td>2,150,000</td>
<td>25,000,000</td>
<td>12,000,000</td>
<td>2,350,000</td>
<td>2,300,000</td>
<td>4,500,000</td>
</tr>
<tr>
<td>Total effort (person months)</td>
<td>281</td>
<td>9241</td>
<td>2319</td>
<td>277</td>
<td>340</td>
<td>639</td>
</tr>
</tbody>
</table>

6.1 EUROPEAN GRID INFRASTRUCTURE – INTEGRATED SUSTAINED PAN-EUROPEAN INFRASTRUCTURE FOR RESEARCHERS IN EUROPE (EGI-INSPRE)

EGI-InSPIRE will support the establishment of a sustainable model for a European Grid Infrastructure (EGI) that integrates resources contributed by national and domain-specific resource providers. Key to this process is a new organisation, EGI.eu, coordinator on behalf of the European resource provider community of the EGI-InSPIRE project, which is also more broadly dedicated to coordinating the EGI community on behalf of its stakeholders.

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

The production infrastructure supports Virtual Research Communities – structured international user communities – that are grouped by specific research domains. Virtual Research Communities are formally represented within EGI at both a technical and strategic level. Direct support is coordinated through a central helpdesk, that brings together operational, technology and other support teams from within the EGI-InSPIRE project and other partner projects.

The EGI-InSPIRE project focuses principally on the European production infrastructure, it needs to support the collaborative research needs of its user communities, for their resources to be integrated with infrastructures around the world. In addition to over 40 partners located within geographical Europe, EGI-InSPIRE includes 8 unfunded partners from the Asia Pacific region. Strong collaborations are also expected with infrastructures in North America, and the emerging infrastructures in Latin and South America. EGI-InSPIRE will support and develop the European DCI community in three important ways:

- Integrate resource providers within the National Grid Initiatives (NGIs) and European International Research Organisations (EIROs).
- Support the development of policies to ensure effective technical management, integration and operation of the EGI for its user communities.
- Coordinate the development and support of structured Virtual Research Communities currently using the production infrastructure within the European Research Area.
EGI-InSPIRE’s will provide services to the community through the partners within the project to:

- Operate a secure, integrated, reliable pan-European infrastructure that can support diverse science communities through the deployment of different technology solutions.

- Work with external technology providers (initially EMI and IGE) to ensure that their solutions meet the needs of the operational and user community in terms of reliability, scalability and functionality.

- Provide support to the communities that rely heavily on the infrastructure by supporting the shared services and tools that are common to many of them

Together these services will support a virtuous feedback circle – starting with a set of integrated services on the production infrastructure that meet the needs of its users, working with external technology providers to define new or improved services based on these existing services, assessing the quality of the new delivered services, followed by their deployment into the production infrastructure.

Additional effort within the project will develop the operational tools to fully devolve these to a national rather than a central operational model, while ensuring that resources such as HPC, desktop grids and virtualised resources are fully integrated into the monitoring and accounting infrastructure.

6.2 EUROPEAN MIDDLEWARE INITIATIVE (EMI)

The European Middleware Initiative is a close collaboration of the three major middleware providers, ARC, gLite and UNICORE, and other software providers. It will deliver a consolidated set of middleware components for deployment in EGI (as part of the Unified Middleware Distribution or UMD), PRACE and other DCIs, extend the interoperability and integration between grids and other computing infrastructures, strengthen the reliability and manageability of the services and establish a sustainable model to support, harmonise and evolve the middleware, ensuring it responds effectively to the requirements of the scientific communities relying on it.

European scientific research has benefited in the past several years from the increasing availability of computing and data infrastructures that have provided unprecedented capabilities for large scale distributed scientific initiatives. A number of major projects and endeavours, like EGEE, DEISA, WLCG, NDGF, OSG, See-Grid, BalticGrid and others, have been established within Europe and internationally to share the ever-growing amount of computational and storage resources. This collaborative effort has involved hundreds of participating research organizations, academic institutes and commercial companies. The major outcome is a number of active production infrastructures providing services to many research communities, such as High Energy Physics, Life Sciences, Material Science, Astronomy, Computational Chemistry, Environmental Science, Humanities and more.

At the core of these rich infrastructural facilities lies the grid middleware, a set of High Throughput Computing (HTC) and High Performance Computing (HPC) software services and components that enable the users to access the distributed computing and data resources, execute jobs, collect results and share information. Middleware like gLite from the EGEE project, ARC from the NorduGrid Collaboration, UNICORE, VDT, Globus and other specific services for computing and data management have allowed thousands of scientific researchers to access grid-enabled resources and produce scientific results.

After the necessary initial period of research and consolidation that took place in the past 6 to 8 years, the growing usage of distributed computing and data resources by scientific communities and individual researchers requires now the stabilization of the computing infrastructures and a
simplification and standardization in the use of the associated software tools. It is of strategic importance towards the establishment of permanent, sustainable research infrastructures to lower the barriers that still prevent potential communities of tens of thousands of scientists and researchers to consider grids as a commodity tool serving their daily research activities. The ultimate vision is that establishing distributed scientific collaborations and using distributed computing and data resources should be as easy as opening a web application, entering simple identification information, entering a few clear parameters to define the task to be executed and its requirements and then waiting for the results to be made available in a well known, easily accessible place.

The EMI project will make the realization of this vision possible by addressing and solving a number of problems that today still prevent users from easily accessing and using the existing computing infrastructures:

- Usability will be enhanced by removing redundancy and consolidating the services, simplifying the security management without compromising its strengths, adding integrated support for high level gateways and portals and transparently making use of virtualization to increase resource availability and management.

- Compatibility will be improved by removing proprietary interfaces in the middleware services and ensuring true interoperability through the adoption of agreed community standards.

- Manageability will be improved by providing standard service configuration, monitoring and instrumentation interfaces and making accounting and other operational information more readily accessible.

- Interoperability between grids, supercomputers and emerging computing models like clouds and desktop grids will be extended to address scalability and accessibility requirements.

- Sustainability will be improved by establishing collaboration programs with commercial companies, adopting off-the-shelf components to reduce maintenance costs and to facilitate easier adoption by wider user communities. The definition together with the resource providers of measureable Service Level Agreements will provide the base for establishing more standard service provision business models.

6.3 INITIATIVE FOR GLOBUS IN EUROPE (IGE)

The Initiative for Globus in Europe, IGE, serves as a comprehensive service provider for the European e-infrastructures regarding the development, customisation, provisioning, support, and maintenance of components of the Globus Toolkit, in close collaboration with the European Grid Initiative (EGI), Distributed Computing Infrastructure (DCI) projects, and Standard Development Organisations (SDOs).

By coordinating the European Globus activities, IGE drives forward Globus developments according to the requirements of European users and strengthen the influence of European developers in the Globus Alliance. This strengthens the representation of European topics such as security and privacy, data privacy protection, compatibility with Grid standards used in Europe to enable interoperability, and aspects of multi-nationality within Globus and the Globus Alliance.

6.3.1 Objectives

The overarching objective is to help the European researchers by lessening their hassle with using DCIs and allowing them to harness greater computing power already available (such as DEISA or PRACE). More specifically, IGE
DNA1.4 - EMI ROADMAP AND DCI COLLABORATIONS

Doc. Identifier: EMI-DNA1.4-1277542-EMI_Roadmap_DCI_Collaborations-v1.0.doc
Date: 30/09/2010

- Adapts Globus to better fulfill European requirements by coordinating European input from both users, developers, and infrastructure providers and thereby strongly impact the open source progress of the Globus Toolkit,

- Adds the European perspective to Globus by delivering tailored software development, operation, support, training, and documentation services to the European communities, and act as a central hub for Globus within Europe, and

- Broadens the adoption of Globus in Europe through coordinated dissemination, standardisation, and test infrastructure operation to foster seamless use of Grid infrastructures in other parts of the world.

6.3.2 Action plan

Over the past years, Europe has heavily invested in building e-Infrastructure for science. Especially in the area of DCIs, the Globus Toolkit is widely adopted as middleware solution and many scientific communities already contributed large efforts into using their application on top of the Globus middleware. Therefore, it is crucial to protect these investments during and after the transition to EGI.

To this end, IGE connects the European efforts on Globus usage, development, and operation by providing a single focal point: Through the European Globus Hub, major stakeholders will be able to learn about and use Globus, get involved in development and training, and contribute to the overarching goal of advancing Globus according to European needs. This includes various areas of concern:

Networking: IGE aggregates, consolidates, and provisions experiences in usage, development, and training from European Grid communities with the European Globus Hub. The visibility, presence, and adoption in Europe and maintaining close cooperation with the international community—including the Globus Alliance—is increased through coordinated efforts of the European Globus Liaison Office.

Services: IGE supports the definition and implementation of Grid infrastructures on the basis of Globus Toolkit. To this end, a comprehensive reference installation and test environment is provided which specifically caters the needs of European user, developer, and infrastructure provider communities.

Research: IGE delivers Globus components and tools with a particular European focus which fulfill the specific needs of the European e-Infrastructures’ user communities. In this context, the provision of components that are eligible for inclusion into the UMD provided by EGI and interoperable with other middlewares through agreed standards (e.g. OGSA-BES and JSDL) is paramount. In addition, IGE collaborates with other Globus developers to foster reliability, usability, and stability of the Globus and cooperates with the Globus Alliance to add missing functionality, increase the manageability, and to introduce improvements into the core distribution.

User Integration: The substantial demand for Globus in Europe is shown by over thirty active supporters of IGE, ranging from industry, academia, e-Infrastructures, NGIs and international Grid projects. The seamless and progressive transition in services delivered through IGE will deliver a user-friendly, well-integrated Globus distribution and thus present a transparent and cost-effective way forward for current and emerging user communities in the European Research Area.

Internationality: Especially in the Americas and Asia-Pacific area, Globus is often the solution of choice for building Grid infrastructures. In order to support cooperation with international researchers, Globus is provided in a coherent way to European researchers, thus promoting close collaboration and interoperability with already established research infrastructures worldwide. These transatlantic relations in DCI research, development, and operation is further strengthened by the increase in
international collaboration. As a grassroots movement, a reliable link with the Globus development team is established within IGE by the incorporation of University of Chicago as a full partner.

6.4 EUROPEAN DESKTOP GRID INITIATIVE (EDGI)

EDGI (European Desktop Grid Initiative) is aimed at deploying Desktop Grid and Cloud Computing services for the European Grid Initiative (EGI) research user communities that require large-scale distributed computing resources for multi-national projects. In order to achieve this goal EDGI will develop middleware for extending Service Grids (SG) (ARC, gLite, UNICORE) with Desktop Grids (DG) (BOINC, XtremWeb, OurGrid) enhanced by Academic Clouds (Eucalyptus and OpenNebula). Software components of ARC, gLite, UNICORE, BOINC, XWHEP, Attic, 3GBridge, OpenNebula and Eucalyptus will be integrated into a SG → DG → Cloud platform for service provision and as a result EDGI will extend ARC, gLite and UNICORE grids with volunteer and institutional DG systems. In this way, the whole European e-science ecosystem will benefit from Desktop Grid extensions, since parameter sweep applications that run millions of sequential jobs can be directed from the expensive cluster and supercomputer resources to cheap desktop resources.

EDGI will create novel QoS support for the DG systems and will explore new service provision models in order to ensure harmonised DG→Cloud interfaces to ARC, gLite, UNICORE resources. The developed DG→Cloud bridge middleware has the goal to get instantly available additional resources on demand if the application has some QoS requirements that could not be satisfied by the available resources of the Desktop Grid system. New scheduling algorithms will be developed that will be able to take into consideration QoS requirements and will enable a more flexible allocation of task and resources in the Desktop Grid systems.

EDGI will further develop the support for data-intensive applications and not only in the context of gLite but also in the context of ARC- and UNICORE-based Grid systems. The ADICS P2P data management system and its bridge support developed in EDGeS at prototype level will be extended for ARC- and UNICORE-based Grid systems and will be deployed as production service in the EDGI project.

The figure shows the place of Desktop Grids in the well-known pyramid of computational resources for e-Science. At the top are the supercomputers, the large optimized systems located in supercomputer centres. A number of these machines are connected into a supercomputer Grid, pioneered by DEISA and continued by PRACE. For many applications clusters perform just as well, but they are less expensive and easier to manage. There are many more clusters than supercomputers. Clusters can be connected to cluster Grids, managed typically by EGI/NGI’s. At the lowest level are Desktop Grids. Especially when one looks at volunteer desktop Grids, the number of computers can be even large. Desktop grids are suited for a subset of cluster Grid applications. Clouds can fit in at many levels, but are placed outside the computing pyramid, because of their specific function in the EDGI project. The figure also shows how
the different levels of the pyramid are connected, or will be connected by the EDGI or DEGISCO project.

EDGI does work closely together with the DEGISCO project. DEGISCO is a support project that supports extension of the European DCIs into countries outside the European Union, with a focus on Desktop Grids.

EDGI and DEGISCO did start the International Desktop Grid Federation to support Desktop Grid operators and developers for Desktop Grids. Integration of Desktop Grids into the European DCIs is an important goal.

EDGI will support the European Chapter of the International Desktop Grid federation with the aim of advancing and promoting Desktop Grid technology in Europe both by sharing and mutually leveraging experience and technological solutions acquired while independently operating Desktop Grids (such as Ibercivis, SZTAKI Desktop Grid, AlmereGrid, EDGeS@home, and many others). The federation will work at both technical and dissemination level. At technical level, it brings together Desktop Grid administrators who together will provide best practices and common solutions to common problems and share knowledge (instead of independently coming up with different and incompatible solutions to these problems). This forum is also used to disseminate the bridge middleware knowledge among European Desktop Grid system providers including companies. At the dissemination level, the federation is a key player in reaching European citizens to provide Desktop Grid resources that would not be possible for the individual European Desktop Grids alone. The International Desktop Grid federation will significantly contribute to the sustainability of the EDGI production infrastructure created in the project.

6.5 STRATUSLAB

6.5.1 Summary
StratusLab is aimed at service provisioning, networking and research of cloud and virtualization technologies to simplify and optimize the use and operation of existing distributed computing infrastructures like the European Grid Infrastructure (EGI). The project is developing the StratusLab Toolkit, an open source cloud distribution. It incorporates cloud and virtualization innovation into existing grid infrastructures by integrating cloud technologies and services within grid sites. Further, it enriches existing computing infrastructures with “Infrastructure as a Service” (IaaS) cloud-like delivery paradigms.

6.5.2 Objectives
StratusLab brings several benefits to the e-Infrastructure ecosystem, in terms of simplification, added flexibility, increased maintainability, quality, energy efficiency and resilience of the sites. The new StratusLab Toolkit cloud distribution complements existing grid middleware services: the aim is for the cloud layer to be fully transparent to layers above. Existing grid middleware continues to provide the glue to federate the distributed resources and the services for high-level job and data management. StratusLab will help to improve the usability of distributed computing infrastructures, to attract scientific user communities, to appeal equally to industrial users, to keep European research infrastructures at the technological forefront, and to strengthen the know-how in virtualization and cloud computing of European industry.

6.5.3 Action plan
StratusLab will integrate, distribute and maintain a sustainable open-source cloud distribution to bring cloud to existing and new grid sites. The StratusLab toolkit will be composed of existing cutting-edge open source software and the innovative service and cloud management technologies developed in the project. It will also include the required additions to turn the software elements into a production grade
distribution to support production quality and operational systems, as will be demonstrated with the operation of production level grid sites in the project.

StratusLab is a two-phase project. In the first phase, the project will focus on **cloud computing for resource provisioning in grid sites**. This will entail development and integration of the initial StratusLab cloud platform, incorporating the components required for the virtualization of grid sites; and creation of virtual appliances for the scientific application domains in the project.

In the second phase the emphasis will shift towards developing **new cloud-like delivery paradigms in grid sites**. This will build on the first phase, including new IaaS cloud interfaces and support for creation of new virtual appliances, which will be stored in a repository.

Efforts to achieve both goals will start from the beginning of the project: the expectation is that the second goal will be achieved in the longer term.

**Networking activities:** The project’s networking activities have been designed to foster collaboration over the complete spectrum of actors, from project participants, through our targeted user communities, to the ensemble of related European projects.

StratusLab will undertake extensive dissemination activities, targeting the user communities listed below, as well as the general public. Awareness of the project will be achieved through participation in relevant meetings, forums, workshops and conferences. StratusLab aims to publish in relevant journals and magazines. The project will also be active online through web presence. In-depth knowledge transfer will take place through demonstrations and training sessions.

**Service activities:** In order to certify the StratusLab toolkit, the project will deploy and maintain a small yet representative infrastructure. This ‘pre-production’ environment will provide the required platform for deploying incrementally the results of the cloud integration activity, but also provide a test-bed for joint research activities to deploy and test their research results.

The StratusLab infrastructure will also serve as an important platform for assessing the economic impact of cloud technologies in the provision of grid services both in terms of human resources (e.g. for administration and system maintenance) and environmental costs (power consumption, carbon footprint, etc.)

**Joint Research activities:** In StratusLab the research activity consists of very specific and focused actions to achieve the main goal of the project that is to integrate a toolkit for offering cloud and grid services. The research activity will be targeted to extend current grid site management functionality, providing or enhancing tools and components to define and dynamically support service elasticity and SLA-powered scalability, optimize site provisioning, placement heuristics, virtual images management and resource sharing capabilities.

**User communities:** StratusLab benefits a wide variety of users: scientists, software scientists and engineers, community service administrators, system administrators and hardware technicians.

### 6.6 Virtual Multidisciplinary Environments Using Cloud Infrastructures (VENUS-C)

**Goals:** VENUS-C is aimed at developing and deploying a Cloud Computing service for research and industry communities in Europe by offering an industrial-quality service-oriented platform based on virtualisation technologies, with the aim of:

- Creating a platform that enables user applications to leverage cloud computing principles and benefits.
Leveraging the state of the art to bring on board early adopters quickly, incrementally enable interoperability with existing Distributed Computing Infrastructures (DCIs) and push the state of the art where needed to satisfy on-boarding and interoperability.

Creating a sustainable infrastructure that enables cloud computing paradigms for the user communities inside the project and new communities recruited through an Open Call.

**Operation and Services:** The VENUS-C solution is an Open and generic Application Programming Interface (API) at platform level for scientific applications, striving towards interoperable services. The VENUS-C platform will be based on both commercial and open source solutions underpinned by the Engineering data centre, Microsoft through the Windows Azure and its European data centres, and two European High Performance Computing centres: The Royal Institute of Technology (KTH, Sweden) and the Barcelona Supercomputing Center (BSC, Spain). Azure offers a multi-layer solution, including computing and storage power, a development environment and immediate services, together with a wide range of services that can be consumed from either on-premises environments or the Internet. From an Open Source perspective, the Eucalyptus and OpenNebula solutions will be evaluated, while the Emotive middleware for clouds will be offered by the Barcelona Supercomputing Centre, thus demonstrating interoperability and ultimately portability to the VENUS-C users.

Technical challenges addressed include virtualisation, service orientation and digital convergence, which are at the heart of the cloud model, as well as current open issues on interoperability with existing DCIs (e.g. Supercomputers), Data Management, Programming models, Application Security, Monitoring and Accounting, Networking and Network Security.

**Action plan:** In the first 12 months, the project will focus on the delivery of an end-to-end prototype which delivers immediate value to scientific partners: the first release will focus on dynamic job submission and workload dispatch into multiple underlying DCI and cloud providers. Subsequent milestones will enable integration with data management, security and programming models, working in synergy with our scientific users, primarily focusing on the functionality of directly usable application-level. Less-visible infrastructure work will start after the initial delivery of the core platform.

**User scenarios:** VENUS-C draws its strength from a joint co-operation bringing together industrial partners and scientific user communities through an innovative approach in the drive towards world-class research, and competitive edge for the European research community. The infrastructure will be initially tested across four thematic areas comprising seven applications: Biomedicine (integrating widely used tools for Bioinformatics, System Biology and Drug Discovery); Civil Protection and Emergencies (focusing on early fire detection), Civil Engineering (construction information management for environmental compliance), and data for science (Marine Biodiversity). To broaden the scope of the current user scenarios, VENUS-C will co-ordinate an Open Call, which will fund up to twenty new experiments in order to address the advanced needs of user communities, in some instances handling complex workflows and data-intensive scenarios. VENUS-C aims to empower these communities through the easy deployment of end-user services, in order to make e-Infrastructures more widely valuable across a spectrum of research fields without the complexity of existing grids and high up-front costs.

**New Business Scenarios:** An important goal of VENUS-C is to assess new business models as part of the drive to foster the shift away from the use of credit cards on a pay-per-use basis and placing more emphasis on a spirit of entrepreneurship through the involvement of pioneering European enterprises and outreach to clusters and start-ups. The feasibility of different follow-on scenarios, such as Public-Private Partnerships (PPP), integration with pertinent on-going initiatives, or service provision through open tenders will be investigated and will bring on board the value-add of each VENUS-C partner, whether public or private, underpinned by co-operation and synergies on multiple levels.
**Co-operation with external experts**: Provisioning, deployment, sustainable growth and cost-effective investment at EU level will also be addressed by drawing on the advice and insights of a select group of experts in and outside Europe, recruited from the distributed computing and service-oriented technology arena. To this end, VENUS-C co-ordinates an External International Advisory Committee providing timely input on pertinent initiatives in and outside Europe, coupled with insights on technical and business-level developments and the broader, international landscape that will help position VENUS-C as an EC-funded initiative, in this landscape and help support potential integration, partnerships and synergies across the distributed computing arena.

**Co-operation at EU and International level**: VENUS-C partners have an extensive network of relations with other countries and initiatives with which strategic alliances will be established. VENUS-C objectives on the Open and generic APIs for scientific platform can only be fully reached if it succeeds in liaising with any actor in this context. U.S. initiatives like FutureGrids and Magellan, other R&D EU projects like VISION- CLOUD, Contrail, R&D projects on Experimental test-beds, like TEFIS, Bonfire, etc, projects on Scientific Data repositories like the D4Science Ecosystem, as well as initiatives on impact assessment of e-Infrastructure technologies like ERINA.

**VENUS-C in the European Landscape**: VENUS-C aims to broaden inter-disciplinary scientific collaboration in Europe and to address the following issues in the European landscape.