# EUROPEAN MIDDLEWARE INITIATIVE

## COLLABORATION PROGRAMS

### EU DELIVERABLE: DNA2.1.1

| | |
|---|---|
| Document identifier: | **EMI_DNA2.1.1_v0.2.doc** |
| Date: | **10/06/2010** |
| Activity: | **NA2** |
| Lead Partner: | **INFN** |
| Document status: | **DRAFT** |
| Document link: | |

**Abstract:**
This document describes the EMI collaboration programs, their organization and access criteria, the agreements to be signed by participating partners and the expected results of the collaborations. It is updated during the project to report on the results achieved.

**Copyright notice:**

Copyright © Members of the EMI Collaboration, 2010.

See www.eu-emi.eu for details on the copyright holders.

EMI ("European Middleware Initiative") is a project co-funded by the European Commission as an Integrated Infrastructure Initiative within the 7th Framework Programme. EMI began in May 2010 and will run for 36 months.

For more information on EMI, its partners and contributors please see www.eu-emi.eu

You are permitted to copy and distribute, for non-profit purposes, verbatim copies of this document containing this copyright notice. This includes the right to copy this document in whole or in part, but without modification, into other documents if you attach the following reference to the copied elements: "Copyright © Members of the EMI Collaboration 2010. See www.eu-emi.eu for details".

Using this document in a way and/or for purposes not foreseen in the paragraph above requires the prior written permission of the copyright holders.

The information contained in this document represents the views of the copyright holders as of the date such views are published.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE MEMBERS OF THE EGI_DS COLLABORATION, INCLUDING THE COPYRIGHT HOLDERS, OR THE EUROPEAN COMMISSION BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THE INFORMATION CONTAINED IN THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

**Trademarks:** [example from EGEE] EGEE and gLite are registered trademarks held by CERN on behalf of the EGEE collaboration. All rights reserved"

## Delivery Slip

| | Name | Partner / Activity | Date | Signature |
|---|---|---|---|---|
| **From** | Diana Cresti | INFN / NA2 | 10/06/10 | |
| **Reviewed by** | | | | |
| **Approved by** | | | | |

## Document Log

| Issue | Date | Comment | Author / Partner |
|---|---|---|---|
| 1 | 19/05/10 | Table of Contents created | D. Cresti / INFN |
| 2 | 06/06/10 | First draft | D. Cresti / INFN |
| 3 | | | |

## Document Change Record

| Issue | Item | Reason for Change |
|---|---|---|
| 1 | | |

| Issue | Item | Reason for Change |
|---|---|---|
| 2 | | |
| 3 | | |

# TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1. PURPOSE

This document outlines the initial collaboration plans for EMI.

## 1.2. DOCUMENT ORGANISATION

## 1.3. APPLICATION AREA

## 1.4. REFERENCES

**Table 1: Table of References**

| | |
|---|---|
| **R 1** | |
| **R 2** | **EGEE Collabrating Projects**, http://collaborating.eu-egee.org/index.php?id=165 |
| **R 3** | **EMBL – European Molecular Biology Laboratory**, http://www.embl.fr, http://www.embl-hamburg.de, http://www.embl.de, http://www.ebi.ac.uk/, http://www.embl.it |
| **R 4** | **Keeping an eye on the skies with LifeWatch, poster winner of EGEE 09** (iSGTW feature article), http://www.isgtw.org/?pid=1002239 |
| **R 5** | **EGEE RESPECT program**, http://technical.eu-egee.org/index.php?id=290 |
| **R 6** | **Opinion: Africa Grid?** (iSGTW article), http://www.isgtw.org/?pid=1002403 |
| | |
| | |

## 1.5. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to XXX. The procedures documented in XXX have been followed.

## 1.6. TERMINOLOGY

**Table 2: Table of Definitions**

| | | |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| | | |
|---|---|---|
| | | |

## 2. EXECUTIVE SUMMARY

EMI will be involved in a large number of collaborations within Europe and globally. The nature of these collaborations is fairly heterogeneous, with different strategies employed depending on the objectives of the collaborating party vis à vis the middleware services provided by EMI.

This document provides a rough classification of the kinds of collaborations envisaged at month 1 of the project. This categorisation is applied as a matter of convenience, and a number of collaborating projects can easily fit into more than one category; furthermore, even within a category the specific modes of collaboration may vary considerably. Ultimately each collaboration will be managed according to its specific characteristics, to ensure optimal synergies in each case.

This document is not intended to provide a full list of current or planned collaborating projects; such a list, together with relevant details on what relationship each project or party has with EMI, will be published on the EMI website, and of course kept regularly updated.

# 3. EMI COLLABORATION PLAN AT MONTH 1 OF THE PROJECT

At this early stage in the project, the EMI collaboration plan must be considered as a statement of intention, more than a programmatic document. Collaboration requires agreements with other parties on specific actions and ways of optimising services which EMI cannot in principle predict.

In particular, because the EGI and PRACE ecosystems are not fully in place at the time of writing this document, all considerations in this section below must be considered very preliminary.

NA2 is in discussions with its counterparts in some of these projects to optimize information sharing …

## 3.1. INFRASTRUCTURE, STANDARDISATION AND PRIVATE SECTOR COLLABORATIONS

This level of collaboration includes direct participation of project members to specific activities by the DCIs and standardization bodies, or dedicated technical workshops to produce some kind of results like reports or recommendations or advancements of the state of the art.

### 3.1.1 Distributed Computing Infrastructures

The collaborations with EMIs main stakeholders – EGI, PRACE, NorduGrid, etc. – will be particularly strong across all the project activities. With a few exceptions, the co-awareness in this area is already leading to joint activities such as the "Engaging European DCIs Together" workshop organised by the SIENA project (http://www.sienainitiative.eu/) in May-June 2010.

#### 3.1.1.1 The Large Customers

**EGI**, the European Grid Initiative (http://www.egi.eu) is a federation of independent National and domain specific resource providers, who support research communities both within Europe and worldwide. EGI.eu brings together European partner institutions – in particular the National Grid Initiatives (NGIs) – to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the user community. EGI is funded in part by the NGIs themselves and in part via the EC project EGI-InSPIRE.

**PRACE**, the Partnership for Advanced Computing in Europe (http://www.prace-project.eu/) prepares the creation of a persistent pan-European HPC service, consisting of several tier-0 centres providing European researchers with access to capability computers and forming the top level of the European HPC ecosystem. PRACE is a project funded in part by the EU's 7th Framework Programme.

**NDGF**, the Nordic DataGrid Facility (http://www.ndgf.org) is a collaboration between the Nordic contries (Denmark, Finland, Norway, Sweden). NDGF already works closely with EGI.

EMI's relationship with EGI is expressed on several different levels, partially described in the following sections. The collaboration with NDGF is ensured by the EGI relationship on the one hand and the co-presence of ARC partners on the other.

The collaboration with PRACE is in part reflected in the UNICORE services, and in part by recent liaison activities under earlier projects (EGEE III and EGI_DS); further work will be necessary to enhace this collaboration.

#### 3.1.1.2 DCI Technology Partners

**IGE**, the Initiative for Globus in Europe (http://www.ige-project.eu/) provides the European link with Globus technology. A close collaboration with IGE will ensure better interoperability and convergence on standards.

**EDGI**, the European Desktop Grid Initiative (http://www.edgi-project.eu) will develop bridge middleware that integrates ARC-, gLite- and Unicore-based Grids with Desktop Grids (BOINC and XtremWebHEP-E). Interoperability of grid services with desktop grid technology already began with gLite during the EGEE III project. The EMI collaboration with EDGI will further extend the to ARC and UNICORE, and in general all the middleware supported by EMI.

**StratusLab** (http://www.StratusLab.eu) will incorporate virtualisation and cloud technologies into existing and future grid infrastructures. The StratusLab Toolkit will benefit infrastructure operators through simplified administration, increased flexibility, and improved resilience. For research communities, StratusLab will improve the usability of distributed computing resources, attracting scientific and industrial users who have embraced the cloud computing model. It will also strengthen European know-how in virtualisation and cloud technologies.

**Venus-C** (http://www.venus-c.eu/) will develop and deploy a Cloud Computing service for research and industry communities in Europe. This will offer an industrial-quality service oriented platform based on virtualisation technologies. The project aims to enable easy deployment of end-user services, thus making e-Infrastructures more widely useful to a range of research fields. The project will be industry led and will have access to Microsoft Azure, a major commercial Cloud service.

The integration of core middleware services such as those provided by EMI and Globus with Desktop Grids and Cloud-related toolkits will establish a new level of pervasiveness for distributed computing. The projects must take great care in sharing the relevant tools and information in a transparent and easily accessible manner. NA2 will work in this context to make sure the technical activities remain well connected and do not lose sight of progress being made on all sides.

### 3.1.2 Standardisation and Networks

The standardisation work within the EMI project and with other technology partners such as OGF, the International Grid Trust Federation (IGTF), the Virtual Data Toolkit (VDT), etc. is of course one of the main activities in JRA1, which will produce specific detailed deliverables on all such activities. There are also collaborating projects which support standardisation meetings and events, and . such as the SIENA project (http://www.sienainitiative.eu/). NA2 will maintain close ties with SIENA and similar initiatives to ensure that information about meetings, activities and results is properly communicated among collaborating partners and to the general public.

The relationship with network providers is of course essential to the project. In particular, the crucial aspect of accessibility and network security must be coordinated with some key non-Grid projects like GEANT3 and TERENA.

### 3.1.3 Private Sector Collaborations

Collaborations with commercial partners on specific topics are included in this category. Discussions begun during the preparation of the EMI proposal with **Google** and **Red Hat** are currently being taken up more programmatically. The project has signed MoUs with these partners focusing on two types of activity:

1. Assessment of technology provided by the software vendor that could be integrated in the EMI stack eventually replacing other components and of technology provided by EMI partners that

could be integrated in commercial offerings or applications. Development of an integration roadmap, possibly with the implementation of prototypes.

2. Assessment of the market and targets for commercial usage of middleware or purchase of service provision and maintenance contracts.

The activities with Google focus on the integration of Google cloud technology with the EMI grid service, and assess possible pay-per-use business models for these integrated services. The activities with Red Hat focus on the assessment of the MRG messaging technology, its suitability as messaging framework for the EMI middleware, its enhancement with security technology from EMI and its reintegration in Fedora or Red Hat supported distributions

Another aspect of the Google collaboration, which will be taken up by NA2, will be the work to create and conduct professional usability studies, and to conduct in-reach activities with the EMI developers to foster a culture of user awareness within the project.

The NA2 task TNA2.5 will be particularly active in this area, and will report on progress made by EMI's activities with commercial partners and lessons learned / actions taken towards the goal of increased usability. TNA2.3 will also be active in the context of the in-reach activities.

## 3.2. USER COMMUNITIES AND RELATED PROJECTS: THEMATIC AND REGIONAL COLLABORATIONS

EMI will inherit all current collaborations with communities served by the Middleware consortia, whether they be in the form of projects, Virtual Research Communities, Associations, or simple groups of users.

EMI will also work with new communities and with the large ESFRI infrastructures to provide distributed computing services for an increasing user base in the European Research Area.

Several NA2 tasks have specific plans to organise events (Middleware Forum, meet-the-developers day), conduct usability studies…

### 3.2.1 Continuing Collaborations

EGEE III had a large group of collaborating projects, of which a summary can be seen at [**R 2**]. [similar info for ARC, UNICORE, dCache]. Of these, some have come to an end and others are continuing; others still are entering a new phase, and may have applied for a new cycle of funding.

It is not known at this time exactly which projects are currently active. Nevertheless, it is clearly the case that the communities reached by these efforts, who are now using DCIs on a regular basis, must continue to be served and must have a voice in the evolution of the middleware.

For instance, **D4Science** (http://www.d4science.eu/) provides interoperation among data and digital repository e-Infrastructures, creating the core of an e-Infrastructure Ecosystem on a global scale. The project has collaborated with EGEE since its initial phase as the DILIGENT project, which began in 2004. Its infrastructure serves a number of large clients, including biodiversity scientists, fisheries and aquaculture resources management scientists and decision makers, ICIS, FishBase, SeaLifeBase, GBIF. It is a natural liaison for many data management issues of interest to the EFRI projects. Finally, it continues to work with other EGEE collaborating projects, such as GENESI-DR.

**e-NMR** (http://www.enmr.eu/), now in its second phase (under negotiation), is deploying and unifying the NMR computational infrastructure in system biology. The project will provide the biomolecular NMR user community with a platform integrating and streamlining the computational approaches necessary for NMR data analysis and structural modelling. Access to the e-NMR infrastructure is provided through a portal integrating commonly used NMR software and grid technology. It is

expected that the e-NMR portal will serve a large biomedical community, and the grid services linked to this portal must continue to serve this community efficiently in the foreseeable future.

**GENESI-DR** (http://www.genesi-dr.eu/) works on establishing open Earth Science Digital Repository access for European and global science users. The project builds upon the existing operational and focused Earth Observation (EO) European infrastructure and involves key Earth Science centres responsible for operational data acquisition, processing, archiving and distribution. GENESI-DR works in collaboration with D4Science as well as other projects, some of which have come to an end. Many Earth Science communities rely on the work of this project; these communities – via the **DEGREE** project (http://www.eu-degree.eu/) – have generated some of the best known user requirements that EMI is working to satisfy, e.g. MPI technology.

For each of these projects, EMI will make specific agreements, typically sealed with the signing of a Memorandum of Understanding, to carry forward the services already in use, the continuing collaborations, and the exchange of feedback concerning the quality of services for the user communities.

For user communities not served by a funded project, EMI will make specific targeted efforts to ensure communication and continuity of service. Some travel money has been allocated to unfunded parties for these purposes.

Another important set of collaborations is with a number of projects and activities working on dissemination. Some examples are the **GridTalk** project (http://www.gridtalk-project.eu/), **iSGTW** (http://www.isgtw.org/), the **Belief** project (http://www.beliefproject.org/). These collaborations will be taken up in particular by NA2.2, which is already in the process of planning a possible inter-project "SuperPR" group.

### 3.2.2 The Large User Communities: VRCs and ESFRIs

Some of the current large user communities have been provided a degree of support as part of the EGI-InSPIRE project. For these, EMI will have several channels of communication due to the relationship with EGI-InSPIRE, in particular via the Middleware Coordination Board.

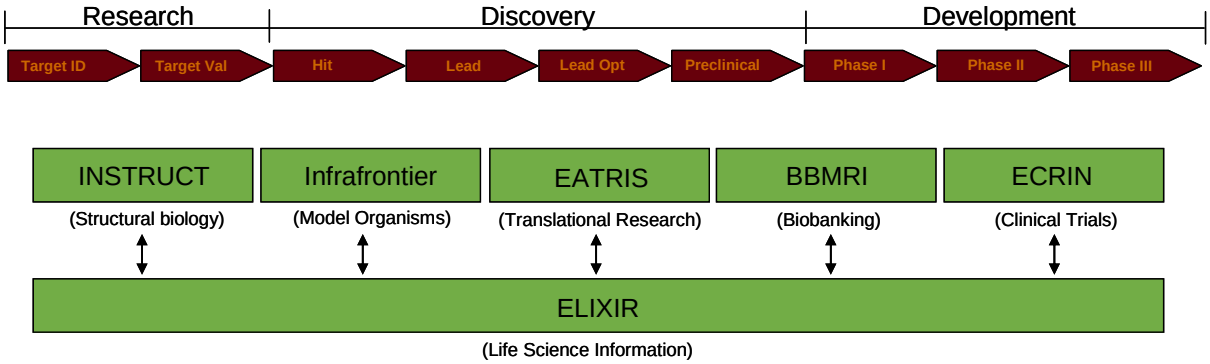
A number of projects related to the EC "Virtual Research Communities" call should become known in the near future. EMI will establish ties with all these projects and, depending on their specific descriptions and objectives, set up appropriate collaboration plans with them.

Finally, the work with the ESFRI projects that was initiated under EGEE III will be taken up in a coordinated manner with EGI-InSPIRE as well as via targeted efforts, which could take the form of specific usability studies, possible "bridge" projects, or other collaborations with relevant projects.

One opportunity that immediately presents itself is provided by the presence of EMBL [**R 3**] in the EGI Council. The EIRO's European Bioinformatics Institute (EMBL-EBI) in Hinxton, UK, will serve as a hub for the **ELIXIR** Bioinformatics Infrastructure, which will also serve several other ESFRI infrastructures as shown in the diagram below:

**Figure 1: The ELIXIR Infrastructure** *(Image courtesy of Andrew Lyall, ELIXIR)*

| | | |
|---|---|---|
| **INSTRUCT** | Integrated Structural Biology Infrastructure | www.instruct-fp7.eu |
| **Infrafrontier** | Infrastructure for Phenomefrontier and Archivefrontier | www.infrafrontier.eu |
| **EATRIS** | The European Advanced Translational Research Infrastructure | www.eatris.eu/ |
| **BBMRI** | European Biobanking And Biomolecular Resources | www.biobanks.eu |
| **ECRIN** | Infrastructures For Clinical Trials And Biotherapy | www.ecrin.org |
| **ELIXIR** | Upgrade Of European Bioinformatics Infrastructure | www.elixir-europe.org |




Preliminary discussions have suggested that this infrastructure may benefit from a set of grid services, possibly (initially) handled via the BioMed VO.

Other discussions have taken place with **LifeWatch** (http://www.lifewatch.eu/), the infrastructure for biodiversity data and observatories. The project has participated in the most recent EGEE events, and its FlySafe application has been featured in iSGTW [**R 4**]. A joint effort has collected a few use cases, around which to base a collaborative "bridge" project. Whether or not such a project emerges, the collaboration with LifeWatch will continue and EMI will play its part.

Further efforts of this kind are under way with other ESFRI projects. This document (or its successor) will be updated when more information is available. For real time updates, please see the EMI website.

### 3.2.3 Collaborations with other World Regions

In the USA, the **Open Science Grid** (OSG, http://www.opensciencegrid.org/), which supports components of the VDT, has had a long and fruitful collaboration with EGEE and the WLCG. This collaboration is continuing seamlessly through various pre-established contact points, most notably the OGF Production Grid Infrastructure Working Group (PGI-WG, http://forge.ogf.org/sf/projects/pgi-wg).

In the Asian continent, EMI has specific strong ties with South Korea and Taiwan, thanks to the EMI partners KISTI and ASGC (respectively). In addition, a number of EMI partners are also partners in the **EUAsiaGrid** (http://www.euasiagrid.org/) and **EU-IndiaGrid2** (http://ww.euindiagrid.eu) projects.

Similar ties with the Latin American continent are provided via the **EELA-2** project (http://www.eu-eela.eu/), and a number of initiatives in the Mediterranean (e.g. **EUMedGrid Support**, http://www.eumedgrid.eu/) and now sub-Saharan Africa [**R 6**] nearly complete the picture of global collaboration.

Coordinating the regional collaborations, and supporting the relationship of EGI with non-European NGIs is a new project, **CHAIN** (under negotiation).

### 3.3. WORKS WITH EMI

As EMI's user base increases, it continues to diversify and become less specialized in distributed computing technologies. Thus EMI services are increasingly accessed through portals and other high level tools, or directly by applications.

"Works with EMI" is a managed collaboration program based on Memoranda of Understanding similar to the EGEE RESPECT program [**R 5**], but more engaging. This program is dedicated to application providers and software developers who need to be aware in advance of internal technical details in order to develop their applications using the EMI APIs or relying on the EMI functionality. This program, which NA2 will coordinate, will also give access to the EMI quality assurance and certification tools and if required to the EMI repository as a means to store and distribute the software products.

All the EGEE RESPECT tools will be included in the program. In addition, … UNICORE and ARC – related tools

The Works-with-EMI program will complement a similar EGI-InSPIRE activity (details added later), but from a more hands-on perspective. The two activities must be coordinated to some extent to avoid duplication of work and dispersal of resources.

## 4. CONCLUSIONS

At month 1 of the EMI project, all collaboration plans can be outlined mostly from the perspective of what EMI intends to do. The plan described in this document should thus be understood as a work in progress, which will be updated as the parties mentioned begin active communications with EMI, and others emerge.