EUROPEAN MIDDLEWARE INITIATIVE

EMI ARCHITECTURE AND TECHNOLOGY DEVELOPMENT PLAN

EMI DOCUMENT

Document identifier:	EMI-DOC-NA1-Technology_Development_Plan-v1.0
Date:	30/04/2013
Activity:	NA1, JRA1
Lead Partner:	LU, Project Technical Board
Document status:	Final
Document link:	https://cds.cern.ch/record/1551298

Abstract:
This document provides a brief overview of the EMI architecture and the technology development directions presented by the four EMI technology areas and by EMI partners. The report represents the final revision of EMI technology planning covering a time period beyond the project end.

TABLE OF CONTENTS

1. INTRODUCTION

The European Middleware Initiative (EMI), a collaboration of the major European middleware providers ARC, dCache, gLite and UNICORE, delivered a consolidated set of middleware components for deployment in EGI, PRACE and other Distributed Computing Infrastructures. EMI fulfilling user requirements extended the interoperability between grids and other computing infrastructures and strengthened the reliability of the services. The EMI project carried out its work along the four technical areas of compute, data, security and infrastructure. The compute area focused on consolidation of standards and agreements through a unified interface for job submission and management, a common format for accounting, the wide adoption of GLUE schema version 2.0 and the provision of a common framework for the execution of parallel jobs. The security area worked towards a unified security model and to lowering the barriers to Grid usage by allowing users to gain access with their own credentials. The data area targeted implementing standards to ensure interoperability with other grids and industry components and to reuse already existing clients in operating systems and open source distributions. The highlights of the infrastructure area were the consolidation of the information system services via the creation of a common information backbone.

This high-level overview paper lays down the future software development directions of the EMI products of the four technical areas after the end of the EMI project. With this document it is emphasized that the evolution of the EMI software ecosystem will continue way after the 30 April 2013 date, the last day of the EMI project. The document is a result of common effort of all EMI partners, EMI Project Office and the project technical management (PTB) to address the review recommendation of producing a post-EMI technology development plan: the EMI project office carried out interviews with Product Team leaders and Partner representatives regarding their future plans while the area leaders provided an overall area view.

In what follows, the technology development plan for the EMI portfolio is presented. First a brief note on grid architectures is given, and then the foreseen future directions within the four areas are described. Finally, the detailed account on the individual partner development plans is completing the report.

2. A NOTE ON GRID ARCHITECTURE(S)

Attempts to define the architecture for "the grid" go back to the early days of the Open Grid Forum. The Open Grid Service Architecture (OGSA) was a conceptual step forward in the days of monolithic systems with proprietary interfaces, and as such motivated new architectural approaches in middleware design. EMI was ultimately able to benefit from this concept, since the contributors came with already distinct modularized components, and the remaining aspect to be fixed were common interfaces. It soon emerged that individual modules, even with identical interfaces, don't have a tendency to self-organise into a coherent infrastructure, especially so when key binding capacities are not provided by any service.

The flagship working group of the OGF, after years of discussions, hundreds of meetings and conferences, use case collection [R1] and terminology definition [R2] published the second revision of their architecture document, "The Open Grid Services Architecture, Version 1.5" [R3]. Several experts from the EMI consortia partners took active role in the OGF architecture process. The v1.5 grand architecture document from 2006 was meant to be embraced as the blueprint for standard-based grid computing. Something that never happened: the implementations and the deployed production grid services rendered the OGSA architecture irrelevant.

The core ideas from the Open Grid Service Architecture, namely that the grid is structured around a coherent set of key capabilities, such as *Execution Management*, *Data*, *Resource Management*, *Security*, *Self-Management* and *Information;* the capabilities are realized by loosely coupled services; and the capabilities themselves are not monolithic and it is possible that only a subset of a given OGSA capability may be present, were picked up by the EGI project in 2010 when there was another attempt made to classify the grid components to be released under the Unified Middleware Distribution. EMI technical management and partners, through the EGI-TCB channels, contributed considerable effort for this yet-another architecture drawing process that run for almost a year. As of Today, the outcome, the UMD architecture [R4] view and the UMD capabilities, are almost completely forgotten, hardly used by anybody.

After all these failed architecture attempts EMI has taken a more pragmatic way and does not try to force an artificial complex structure, a grandiose architecture on its software offerings. Just like nobody talks about a Redhat or Debian architecture, similarly, the EMI software distribution, or the "EMI architecture" should be considered as a set of features and services released together that tackles a set of requirements and can handle a set of use cases. We also note that the term service is used here in a broader sense than (web) service.

Nevertheless, it is possible to group the EMI services into different categories based on their role played in the grid ecosystem. Such a classification was presented in the EMI Security Architecture document [R5] and a shortened version is inserted below. This grouping can be viewed as the "EMI architecture".

2.1. THE EMI "ARCHITECTURE"

The EMI "architecture" should be considered as a set of features and services released together that tackles a set of requirements and can handle a set of use cases. We also note that the term service is used here in a broader sense than (web) service.

In Grid environments users can be organized in so called *virtual organizations* (VO). They allow the management of users across different institutions without the need to observe intra-institutional management structures and policies.

Sites - on the other hand correspond to local installations of computing resources (clusters, storage, etc). Agreements are made, between the individual sites and VOs, that give the members of a VO access to the resources of the sites. In order to do so, sites must install certain services from the middleware stack chosen by the VO. It is important to note that the support of VOs by sites does not mean that the site hands over some of its autonomy, particularly in security issues, to VOs. On the contrary, the local site autonomy must be preserved and respected by the VO and its members at all times.

Besides VO and sites, there are also a set of *common services* that act as a glue between VOs and sites. They can be considered as a part of the underlying Grid infrastructure and are typically operated by some of the larger sites on behalf of the infrastructure. Examples are information services, credential stores, data catalogues, and management services for VOs etc.

Figure 1 shows the high level picture of the EMI service components divided into the elements of *VO-, site- and common- services*. The EMI middleware consists of the following software components:

- User Interface UI: Suite of components through which the user interacts with the Grid (submitting of jobs, querying component status, access to data stored on the Grid, etc). The user needs to possess a X.509 certificate in order to interact with the Grid services, because the EMI security model is currently entirely PKI-based. The UI contains clients that interact with both ARC, gLite and UNICORE services.
- VO services: Users are permitted to access the Grid due to their VO membership. Currently; ARC and gLite use the Virtual Organization Management Service (VOMS). UNICORE uses UNICORE Virtual Organization Service (UVOS).
- Common services: These are services that span the Grid and may serve one or more VOs at the same time. Examples are the information system, file catalogues and resource brokers.
- Site-specific software components: These are the Grid services that the individual sites operate. They comprise of Compute Elements (CE) and Storage Element (SE) services. The user can either access these services directly or through resource brokers and/or file catalogues.

In addition to the EMI components, the Architecture overview figure also depicts some important external services that are necessary to operate a functional grid:

- CA: A certificate authority provides the X.509 credential to the user. The CAs are coordinated through the international Grid trust federation (IGTF) and are outside the purview of the EMI security project. Note that CAs also issue host or service certificates to identify hosts and services. Note that the software needed to operate a CA is not provided by EMI.
- Credential services: The special class of services, which act either as:
    - An attribute authority (AA)
    - A proxy certificate store and renewal service (MyProxy).

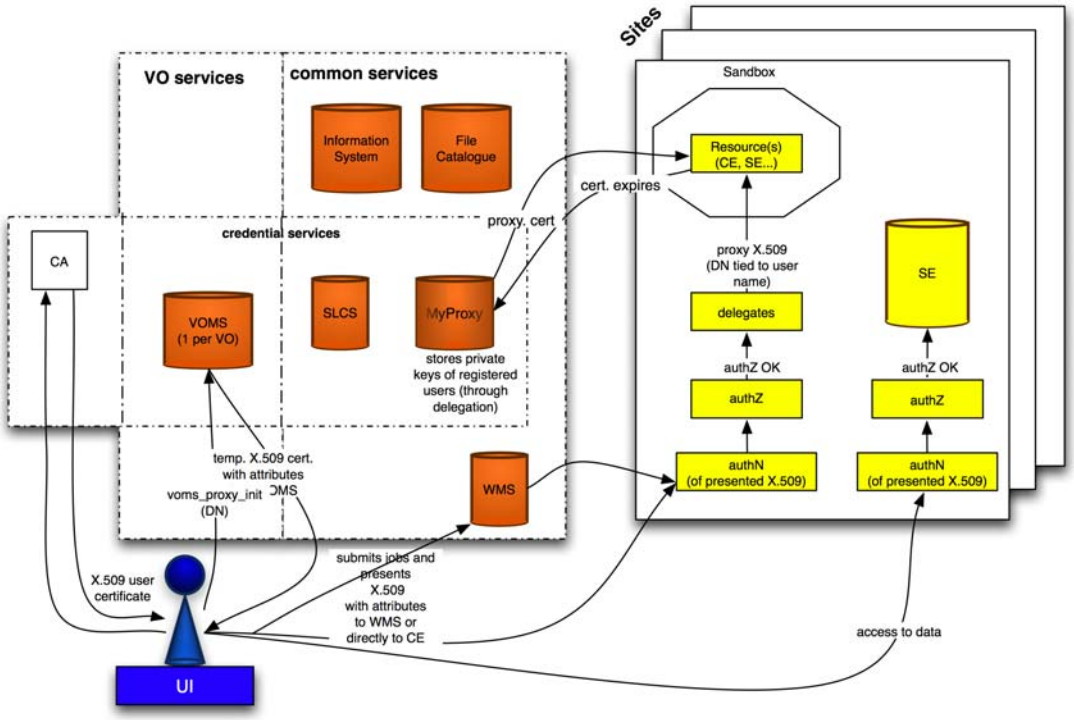

Figure 1 Overview of the EMI architecture composed of common, VO and site services and the corresponding client-side components grouped under the "UI" concept (figure borrowed from [R5]).

Table 2 provides examples for the VO, common, site specific categorization for EMI supported services

Service Class	Offered functionality	EMI supported instance
VO services	VO management system	VOMS and UVOS
	User ID protection	pseudonymity
Common services	Information system	Berkeley Database Information Index (BDII), EMI Registry, EGIIS
	File catalogues	LCG File Catalogue (LFC)
	Store for proxy certificates	MyProxy
	Short-lived (site-integrated) credential services	Security Token Service (STS)
	Resource broker	Workload Management Service (WMS)
	File Transfer Service	File Transfer Service (FTS)
Site-specific services	Compute Element (CE)	CREAM, A-REX, UNICORE CE
	Storage Element (SE)	Disk Pool Manager (DPM), dCache, StoRM, Hydra
	Authorization service	Argus
	User identity switching	gLExec
	Virtual Machine provisioning	Argus-EES, WNODES

Table 2 EMI service categories

3. TECHNOLOGY DEVELOPMENT PLANS: AREA DIRECTIONS

3.1. COMPUTE AREA

Generally speaking, the advent of the Grid and more than ten years of European projects providing technological solutions for distributed computing have led to the development of the consolidated middleware stack that offers a well-understood and mature distributed computing capability.

The High Energy Physics experiments, made of huge federated collaborations, shaped the architecture of the present Grid middleware. This scenario has led to the development of cutting-edge technology on one side, but it also requires high-skills and non-trivial effort to be set up and maintained. Not all the small communities are capable to do that. This simple fact immediately reveals one main direction for the future evolution of the middleware: make it suitable for small/medium sized collaborations as well. To achieve this, the present computing area services must be made able to cope with more dynamic designs.

Reconsidering the role of the present Grid Computing Element, endorsed with its well-established authentication mechanisms, but made potentially able to provide a more general and dynamic access to computing and storage than the classic abstraction based on the concept of *job*, might be a viable option for the near future. Current Grid middleware computing services, in fact, were mostly conceived as interfaces to the traditional batch systems. These, in turn, are constantly evolving, due to changes in the underlying hardware technologies, such as virtualisation, system software, and usage patterns. The CE, in the first place, will have to evolve with the underlying resources and use cases. Inevitable requirements to make efficient use of multi/many-core architectures should result in functionality that is natively built into the middleware, not merely added as an external component. In this respect, the work of EMI to create a *de-facto* and pragmatic standard for the European Compute Elements puts our software in a advantageous position with respect to other, less limited, initiatives.

It happens, in fact, that all the major HEP experiments utilize a centralized framework to gather and process all the computing requests. The framework is kept agnostic of the complexity of the underlying systems in terms of computing paradigm (grid, cloud, volunteer, anything), technology stack, data distribution protocols. Resource provisioning is always decoupled from job management, recurring to the so called 'late-binding' approach in most cases. Various frameworks are nowadays utilised by the major HEP experiments, yet there is no standard handling of security nor a common protocol for the implementation of late-binding. Moreover, they have evolved independently, oftentimes have grown outside the mainstream middleware projects, so that have been developed without complying to well-defined standards. Also, they basically provide the same solution, each with different dependencies and technology stacks. Finally, they are complex in their architecture, and for a small institution wanting to perform accurate workload on its payloads, it would require several time, complex hardware and network configurations, and devoted people to properly set up a VO framework. Under this light, there could be great work to be done towards a consolidation and standardisation of the workload management frameworks. This is not immediately felt as urgent by each single major community, as long as maintenance costs are not immediately visible, but it will surely become a boiling topic in the next few years, also in the context of a world-wide and European economic recession phase.

On the client side, the development of user-friendly interfaces adapted for a variety of popular platforms is a major challenge. Given the multitude and variety of scientific applications, it is unlikely that a single tool suitable for all cases will ever be created. A solution could be the provision of a Grid Software Development Kit, which would simplify and speed up development of science-specific end-usertools. Even in this case, EMI has started work in all these fronts, which creates a solid basis for future developments.

The work on providing more usable and consolidated services has already started, also with interesting achievements, with the EMI project itself. Yet making the software more usable and efficient does not only involve middleware providers, it is a complex activity that requires a tighter integration among technology providers, integrators, all the various DCIs and in particular user communities.

3.2. DATA AREA

As a result of the strong EMI consolidation and development work the EMI-Data software components became indispensable ingredients of large European infrastructures, like EGI and WLCG and some components even contribute to prominent projects within the US, as dCache for the Open Science Grid [R6] and UNICORE for XSEDE [R7].

The work plan of the data area has been continuously adjusted to follow the natural deviation from the original EMI proposal and to accommodate longer term requests going beyond the end of the project. New user requirements and enhanced designs of the different scientific communities were taken into account and the individual data product teams are in good shape to continuously support the key infrastructures and customers long after the end of the EMI project.

The areas of future work related to the above mentioned core infrastructures include the work on FTS3, the HTTP/WebDAV ecosystem, the federated storage framework and the usage of the SRM in UNICORE (for technical details consult the final Data Area deliverable [R8].

In addition to these traditional areas of common interest, the main EMI data product teams will explore the field of *Big Data*:

CERN Data Management provides seamless access to their storage from Cloud systems and vise versa, also runs three weeks demonstration cycles for the backbone of the large scale data distribution infrastructure.

dCache.org is collaborating with the University of Applied Sciences (HTW Berlin) to integrate cloud protocols (e.g. CDMI[R9] and Amazon-S3) into the dCache technology stack

Recently, *dCache.org* and *UNICORE* became part of the German *Large Scale Data Management and Analysis* Project (LSDMA [R10]), which main objectives are to consolidate federated storage and equally important, to simplify authentication to distributed system by building-up a federated identity infrastructure.

The UNICORE team will continue working on future improvements in their data access layer, integrating http and cloud like storage providers thus opening up for Big Data.

3.3. SECURITY AREA

The products and solutions of the EMI security area are mostly built upon the X509 technology. This technology offers reliable, proven and robust solutions but at the same time exhibits inflexibility and complexity. Any major change in the underlying technology would risk the breaking of the existing production systems. Therefore any improvements to the Security stacks would result from requirements of the operators and users of the Grid infrastructures. Experience and requirements have shown that any improvements or changes must not impede or disrupt operations. Therefore the two scenarios for changes are that they are incremental or a complete radical change of direction. Any future projects, large or small, should be driven by the user requirements and be controlled by the user communities.

The process and concept of delegation lies in the heart of the X509-based security. The future of the security layer largely depends on how the delegation mechanism would be implemented and used in the future.

One option is to eliminate the need of delegation via new submission models. Direct submission of pilot jobs to the site computing elements (CEs), rather than WMS, would eliminate the need for one stage of proxy handling (delegation) and renewal. This is the case presently for the CERN LHC experiments but is not generally the case for other EGI Grid communities.

From the security point of view, if a site detects a malicious payload or activity, then it may either:

- Ban the pilot credential, therefore protecting itself against further malicious payloads from the pilot service;
- Ban the payload credential, allowing other pilot-submitted payloads to continue.

By eliminating the WMS step, the site that runs the CE can set its own policy on credential lifetime and banning policy.

If the concept of delegation is still needed, then possibility of alternatives to the delegation step should be explored. Possibly by reducing the use of delegations to only where it is strictly needed. Another path could be SAML adoption within the EMI stack, not just at the interface with the user. This would move the trust from the user credential to the service credential. This proposition would require a large amount of work and a strong request from the user communities to justify such an improvement in the security of Grid infrastructures. The attractions of SAML and ETD are obvious but the reality of maintaining a production Grid infrastructure may rule this out. Another effect that is not well-understood is how scalable this option would be. The main users of the data analysis Grid are just not asking for this feature and probably would not accept the disruption.

If the above proposal is too onerous then the possibility of delegating only some of the users' rights could be explored. Currently it is too complicated for the majority of the Grid users (in gLite/ARC and UNICORE) to determine which attributes (rights) should be selected for the credential to be delegated. The only other option is to automate this selection. Therefore the attributes need to be selected at run-time. Today, this task has not been seriously considered as it has not been requested by users. The perceived difficulty of this task coupled with the lack of user requirements makes the potential return on investment seem low.

The overall ease of the middleware could be improved by full integration of the CANl. This would give standard error messages to users and administrators and allow standard analysis and monitoring tools to be built. Also, with the presence of STS, integration to various portals should be emphasized.

Security assessments of the EMI stack should be planned as part of the release process. Some NGIs are already preforming this work on code that will run in their infrastructures. Code reviews were performed during the EMI project and this work highlighted the labour-intensive nature of this work. A more uniform migration of the security code to public repositories such as github would make it easier for developers to contribute back-patches and examine the code and documentation.

In addition to the X509-related challenges, the emerging adoption of virtualization and cloud technologies pose additional security challenges. By definition, the security of commercial clouds is not something that falls under the domain of the EMI security stack. The very act of deploying services and data to resources that are essentially unknown and outside of the policies developed by Grid projects and VOs renders any services untrusted. The components and services of the EMI Security stacks of course technically can be deployed on virtual machines within a commercial "Cloud" infrastructure. Deployment on a "Private Cloud" i.e. a cluster controlled by virtualization technologies is no different to a deployment on a traditional cluster and would enjoy the same level of guarantees from Grid project and VO policies.

The EMI delegation services could be deployed within a couple of commercial clouds in order to provide a loose federation of their infrastructures. The Argus AuthZ service could be deployed on a Cloud node in order to control access to other nodes that a client may have rented. A Cloud client could therefore shift their data processing or storage based on various metrics.

3.4. INFRASTRUCTURE AREA

The main focus of the infrastructure area work plan during the EMI project was around the consolidation of information system technology of the three stacks and the integration work for a consistent accounting ecosystem.

The information system developments in the EMI project resulted in consolidation and harmonization of the existing components. Three concrete objectives were identified and achieved:

1. Common Information Model (all over the components)
2. Common Service Information Interface (on the resource level
3. Common Service Registry (on the discovery level)

These three fully implemented technologies blocks provide flexible building blocks to create consistent and complete information system deployments. The deployment, combination and roll-out of these modules, however should depend on the infrastructure usage. The particular configuration options are mostly determined by the consumer - side constrains, the need not to break current information consumers. Unfortunately, EMI as a middleware core service provider was not in the position to reliably investigate this aspect. Therefore no best strategy was possible to be proposed. Eventual decision on technology migration is not a responsibility of the time-limited EMI project. EMI itself, of course, would like to see an ideal future evolution where all information is described AND consumed in GLUE2 format, services are publishing fresh service data locally via the ERIS (LDAP) interface and the services themself are discovered from the EMIR service registry. Such a scenario would result in the phase out of the Top-BDII and require not to use GOCDB[R11] as an information discovery service.

The information system is now reasonably mature and the support of GLUE 2.0 represents a major milestone in this area. However, there remains a concern with the accuracy of information published. Work to ensure the validity of information before it is published is required to ensure it is of the highest quality possible. The investigation into the use of messaging did reveal some advantages. One advantage that was not mentioned is the reduction in operational overhead that comes with the simplification of the infrastructure when using enterprise messaging technology. The cost-benefit analysis of migrating the information system to use this technology along with a migration plan would be needed.

The main future focus for EMIR is gaining adoption. It has been demonstrated as an elegant solution to the federated service discovery problem. However, to ensure its success it needs to be adopted by a production infrastructure. The developers should work with the infrastructure providers to realize this goal.

The breakthrough of agreeing on standard accounting records and the accounting transport protocol requires the providers of the tooling and services to work together with the infrastructure providers to solve any interoperability issues. Such a future work needs a strong coordination even beyond the EMI project.

Throughout the project a great deal of work on messaging has been happening in the background. This work is not immediately obvious as it is not an external visible component. The experience gained with messaging technology from the EMI project should be built upon, as enterprise messaging technology is seen as an integration framework for services. The infrastructure area products will continue to utilize this underlying technology.

4. TECHNOLOGY DEVELOPMENT PLAN: PARTNER DIRECTIONS

4.1. PARTNER: CERN

CERN Data Technology Provider: FTS, DPM, LFC, GFAL/lcg_util

Concerning future technical plans, CERN Data TP will continue with the promotion of http ecosystem incubated within EMI. DPM is already supporting this; FTS3 and clients will follow. It is also important to show the integration and performance viability of this ecosystem. DPM will integrate more types of storage, e.g. cloud storage. The current FTS3 pilot implementation will move to production.

After EMI, CERN Data TP proposes to have an agreement among technology providers on how software is released and distributed. To this end, it is strongly in favor of streamlining this process using EPEL.

CERN BDII Technology Provider: Resource, Site and Top BDII

The CERN BDII team is involved in EGI and WLCG. Future development plans are dictated by information system use cases defined by WLCG. The current one-person team, in principle, should be enough to implement future requirements.

4.2. PARTNER: CESGA

Part of CREAM CE Technology Provider: SGE-utils

The ongoing collaboration with LIP and IFCA SGE remains strong and will be leveraged to deliver the required product developments, maintenance and support activities after EMI. There are a small number of EMI requirements related to this product that are still open. This collaboration is taking care of these requirements with a view to close them in the very near future.

4.3. PARTNER: CESNET

NGI_CZ/Metacentrum Technology Provider: Gridsite, proxyrenewal, CANL-C

L&B Technology Provider: L&B

CESNET will continue to explore ELIXIR connection via the CZ NGI after EMI. It will also continue to be actively involved in the EGI Federated Cloud activity, participating in three cloud-related mini projects. Proxy renewal and Gridsite are stable products; these are currently on maintenance mode. There are early indications of cloud-related requirements for Gridsite, but nothing concrete to report at this time. CANL-C is developed in three languages[1], with CESNET maintaining the C version. Exploration of the use of L&B in local setting driven by needs by MetaCentrum is . As the L&B released in EMI 3 Monte Bianco is major release for CESNET, no significant feature requests are expected in the immediate future.

4.4. PARTNER: CINECA

UNICORE Technology Provider: UNICORE Security, UNICORE Container, UNICORE Services, UNICORE Clients, CANL-JAVA

gLite Security Technology Provider: STS

See also UNICORE Consortia

CINECA is planning to integrate UNICORE components with existing data visualization elements already deployed to improve the usability of the final service as compute and visualization

[1] C by CESNET, C++ by ARC, Java by UNICORE

functionalities will be available through the same interface. Data visualization is typical requirement from the industry, with which CINECA has a number of exploitation opportunities. CINECA is also looking at developing intuitive and easy-to-use interfaces based on web technology, fundamental to enable SMEs use computational resources for small-scale well-defined applications. Promotion of UNICORE portal, currently in beta, to become a production service is also planned.

4.5. PARTNER: CSIC

gLite MPI Technology Provider: gLite MPI including mpi-start, MPI_utils

Environment and biodiversity are some of the areas CSIS is currently involved in. CSIC is a member of the FP7 COOPEUS project which aims to strengthen the cooperation between US and the EU in the field of environmental research infrastructures. It is contributing to the project development platforms for biodiversity. A number of commercial links are being explored with a local Siemens branch using mpi-start for simulation and another local SME on environment. More detailed information will become available in summer 2013. CSIC products will continue to be supported after EMI for these communities.

Distributed data management, big data, cloud services, data management from other sciences like astrophysics, support for generic parallel jobs – these are some of the development areas in the pipeline. Requirements from the HADOOP community will also drive the evolution of mpi-start.

4.6. PARTNER: DESY

dCache Technology Provider:dCache

For the next 12 months, the dCache team is focusing on requirements from photon science and Intensity Frontier experiments. Some of the technical areas under review include three tier storage (tape, disk, SSD), federated identity solutions within the LSDMA project, cloud protocol implementation (CDMI, S3) with the HTW Berlin, high speed single data client producers (HPC and Fast Photon Detectors), to name a few. As one of the main storage solution providers of WLCG, dCache is concerned with the technical direction WLCG is taking concerning data. dCache is going for standards and WLCG is focusing on a closed-shop in storage.

4.7. PARTNER: FOM

EMIR Technology Provider: gLite Security (ARGUS-EEES, gLExec-wn, LCAS, LCMAPS, LCMAPS-plugins-c-pep)

In addition to EMI and EGI engagements, FOM has ongoing activities with CLARIN, bioinformatics community in the Netherlands, and networking relations with TERENA, SURFsara and Open Connect. Cooperation with EUDAT is also foreseen. It is worth mentioning that there is an incubation program within FOM for technologies, and that its license is commercial-friendly. All FOM/NIKHEF products will be supported - gLExec, LCAS, LCMAPS, LCMAPS-plugins-c-pep for these communities.

Continued development along the existing product lines and technologies is not likely to advance much or bring significant strategic advantages. New and complementary technologies will be used to expand the user base for scalable multi-domain security solutions with regards to access control tools. Other development paths being explored include integration of web services from different applications and integration of networking technologies from TERENA, SURFsara and Open Connect.

4.8. PARTNER: INFN

CREAM Technology Provider:CREAM-CE

VOMS-StoRM Technology Provider: VOMS, StoRM

WMS Technology Provider: WMS

WNoDeS Technology Provider: WNoDes

In addition to its EGI engagements, INFN is planning to work closely with Italian industrial partners via national funded cloud projects. It is also exploring federated authentication with EUDAT. The CHAIN-REDS project contributes to CREAM dissemination and training. CE-Monitor is used by OSG for many years now, with more than one hundred sites. Within OSG ReSS (Resource Selection Service) project, CREAM is released as part of VDT, a scientific grid information service. CREAM with/instead-of Condor is also being discussed as reference compute solution in OSG. For these communities, INFN will continue to support all of its products using the EGI support channel GGUS.

In terms of development, there are several avenues being explored – VOMS and federated identity integration, evolution of VOMS and WMS to meet the needs of WLCG, evolution of StORM to meet the needs of Italian Tier 1, development of cloud storage interfaces, provide storage solutions using standard interfaces like CDMI, provide an interface to a federated cloud with dynamically allocated resources, provide a platform to migrate applications to the cloud and creating an interface for federated grid-cloud infrastructure. WMS could, if there is interest, be a cloud broker.

4.9. PARTNER: UNICORE CONSORTIA – CINECA, JUELICH, TUD, UWAR

UNICORE Technology Provider: UNICORE Security, UNICORE Container, UNICORE Services, UNICORE Clients, CANL-JAVA

Apart from EGI, UNICORE is used in the PRACE and XSEDE infrastructures. Furthermore, there are many initiatives, projects and other activities where UNICORE is used. For example, German projects (UIMA-HPC, MMM@HPC), in Russia[2] and local deployments such as at the Jülich Supercomputing Centre where UNICORE can be used to access almost 500,000 cores.

In addition, many people download the software and deploy it on their local systems, as can be seen from download statistics and support requests coming in through the SourceForge channels.

The UNICORE team (specifically the partners FZJ and UWAR) will provide support on a best-effort basis (as before EMI), and will react to support requests coming in through SourceForge as well as through infrastructure-specific channels such as GGUS in the case of EGI, or the XSEDE ticket system.

There are detailed plans for further development, for example (a) simplifying end-user experience without the need to handle X.509 certificates, deep(er) integration with existing authentication and authorization infrastructures, (b) providing web-based and mobile clients, (c) strengthening data processing capabilities beyond the traditional batch job model, including enhanced data and metadata management, (d) seamless and secure integration of virtualized and cloud-based systems (both compute and storage).

4.10. PARTNER: JUELICH

EMIR Technology Provider: EMIR, EMIR-SERP

EMIR will be supported after the end of the project using the standard EGI support channel GGUS. Mailing list(s) will be created for support, feature requests and development activities. There are early discussions with (new) communities, e.g. EUDAT, CRISP and NorduGrid, evaluating EMIR as a potential service registry.

EMIR development plans are linked with community requirements. Some of the new features for future EMIR release includes support for new types of services e.g. data oriented services. The EMIR lead developer as OGF GLUE2.0 chair is proposing a new activity on creating a new JSON rendering

[2] e.g. http://supercomputer.susu.ac.ru/en/

of schema. The EMIR team has highlighted the merit of the versatile release management in EMI and development integration and synergies during the project. As a cross-institute product, EMIR will need these facilities after EMI.

4.11. PARTNER: LU

NORDUGRID-ARC Technology Provider: ARC-CE, ARC Clients, ARC gridftp, ARC Core, ARC Infosys, CANL-C++

Various research groups all over the world broadly use ARC. In Sweden, the SNIC infrastructure groups offer ARC-enabled resources. The trouble-reporting tool is NorduGrid Bugzilla; the GGUS-based 1st and 2nd level ticket routing (e.g. the existing ARC SU will accept tickets from the 1st and 2nd level support line of EGI) will continue.

The technical plans are described in the ARC Roadmap[3]. One of the perennial challenges is how to manage changing requirements from user communities, and how these are aligned with current architectures.

4.12. PARTNER: NIIFI

NORDUGRID-ARC Technology Provider: ARC-CE, ARC Clients, ARC gridftp, ARC Core, ARC Infosys, CANL-C++

EMIR Technology Provider: EMIR, EMIR-SERP

Grid related effort at NIIFI is linked exclusively for HPC systems. There are four HPC sites in Hungary interconnected using grid middleware. This infrastructure is currently for Hungarian researchers and organizations.

NIIFI is exploring several expansion scenarios. There are plans to purchase additional HPC facilities to augment the capacity of the current setup. Strengthening links between the industry and academia is high on NIIFI's priorities. To this end, there are ongoing discussions with car manufacturing companies Mercedes and Audi based in the Hungarian towns of Kecskemet and Gyor, respectively. The goal is to establish joint research programmes with higher education establishments in the area, leveraging HPC technologies. Once completed, this will result in six HPC and three storage facilities. NIIFI uses extended MoUs and organizational frameworks to establish such collaborations.

The ARC Roadmap is the reference document for ARC-related technical developments. In addition, NIIFI is also exploring federated identity, management services for Hungarian organizations and big data.

NIIFI highlighted the issue of development and deployment of cross-consortia products like EMIR. The technical coordination for such products provided by EMI during the project will need an equivalent structure after April 2013. As a user, NIIFI believes a high level technical coordination, e.g. technical director for each middleware consortia, streamlines communication between users and technology providers.

4.13. PARTNER: STFC

APEL Technology Provider: APEL

STFC products are deployed in a number of infrastructures – APEL clients (EGI), APEL servers (EGI, OSG), APEL/SSM (EGI, OSG, PRACE). There are discussions with EUDAT on storage accounting. Support for STFC products within EGI is guaranteed for these communities.

Medium term plans include further storage development and clouds.

[3] http://wiki.nordugrid.org/index.php/Roadmap

4.14. PARTNER: SWITCH

ARGUS Technology Provider: ARGUS

SWITCH, in addition to EMI and EGI, is working closely with the Swiss NGI, with several scientific communities in Switzerland and is part of big collaborations GEANT and TERENA. For these communities, SWITCH will provide basic support and maintenance of ARGUS until the end of EGI at least, using the GGUS support channel.

Feature requests for ARGUS that are reasonable and make sense can be considered. Some of the areas being discussed include OpenStack integration and AA for cloud under Swiss university project for at least one year from May 2013.

4.15. PARTNER: TCD

National funding for grid computing has dried up in Ireland. Since summer of 2012, there was no longer funding for a Grid manager post. This has resulted in the closure of the NGI Ireland, withdrawal from EGI involvement and withdrawal from most grid-related activities after EMI.

There will be still some involvement in ER-Flow and SCI-BUS with possible future involvement in cloud proposals, but it is unlikely that TCD can obtain sustained funding to feed effort back into EMI.

4.16. PARTNER: UCPH

NORDUGRID-ARC Technology Provider: all ARC components both on server and client side

UCPH is involved in the European Spallation Source (ESS), a multi-disciplinary research center based on the world's most powerful neutron source. It is also collaborating with the bio informatics and geo physics communities of the university. UCPH will continue to support the NORDUGRID-ARC products, using Bugzilla as the support channel.

The ARC Roadmap describes the future development plans of the NorduGrid-ARC products.

4.17. PARTNER: UIO

NORDUGRID-ARC Technology Provider: ARC-CE, ARC Core, ARC proxy, CANL++

The ARC Roadmap describes the future development plans of the NorduGrid-ARC products.

4.18. PARTNER: UH

gLite Security Technology Provider: STS, Hydra and Pseudonymity

UH is working with the canton of Geneva on pilot implementations of STS, Hydra and Pseudonymity. The group has ongoing links with a number of biomed organizations, e.g. French biomed group, CSC Finland biomed group. There is also discussion with CERN on further development of Hydra; and in using CERN KOJI build system. UH-managed products will continue to be supported using the EGI support channel GGUS for these communities.

Any future developments will depend on user requirements; for instance requirements arising from the WLCG pilot STS activity.

4.19. PARTNER: UPJS

NORDUGRID-ARC Technology Provider: all ARC components both on server and client side

Post-EMI, UPJS is looking at strengthening CERN ties specifically with HEP experiments ALICE and ATLAS. UPJS is also in the process of setting up the recently approved Innovation Park TECHNICOM[4]. It is worth mentioning that EMI and NorduGrid collaboration played an important

[4] http://www.minedu.sk/stvrty-vedecky-park-vyrastie-v-kosiciach/ (in Slovak)

role in securing this project. UPJS will continue to provide support for NORDUGRID-ARC components after EMI, using Bugzilla.

The NORDUGRID-ARC development plans are detailed in the ARC Roadmap. These will be discussed at the next NorduGrid Conference in June 2013. In addition, UPJS is exploring other areas, e.g. university cloud, statistical physics simulation and university lectures on distributed and parallel computing.

4.20. PARTNER: UWAR

UNICORE Technology Provider: UNICORE Security, UNICORE Container, UNICORE Services, UNICORE Clients, CANL-JAVA

See also UNICORE Consortia

In addition to the UNICORE Consortia input, UWAR is going to continue to develop and support the components it has been responsible for as well as other components of UNICORE Technology Provider after the end of the project. In particular, UWAR will continue to support security infrastructure for UNICORE including Common Authentication Library, UVOS and monitoring (NAGIOS probes). UWAR is also supporting a growing user community, mostly from Polish NGI users who access infrastructure resources using UNICORE and gLite services. It is worth mentioning that UWAR is working closely with the life sciences communities in Poland and in Europe to provide dedicated solutions based on EMI components. The availability of rpm builds is important to its community. The support will be provided on a best effort basis. Having said this, there is stable financing for the next two years.

4.21. PARTNER: AS

EMI is the primary grid and cloud infrastructure middleware for the e-Science in Taiwan and Asia. Life science, earth science, environmental changes, HEP, social sciences and long term preservation are the scientific communities being supported. There is Taiwan-Philippines joint project focused on disaster mitigation, in particular tsunami and earthquake monitoring, using AS infrastructure and its grid services. AS will continue to support its services within EGI for these communities.

Distributed computing infrastructure is necessary for e-Science and big data analytics in the foreseen future. Middleware is essential. Without the effort to coordinate the development and quality control of the middleware like EMI, these will have to be done by the (local) institute. Interoperability of resource centres and infrastructures would be a big problem in that case. AS will keep extending the e-Science applications to wider scientific disciplines, not just in Taiwan but also in Asia. Only the reliable middleware would be deployed.

The typical outcome of AS e-Science projects are the web portals and the analysis framework (computing model and workflow) over the DCI. Besides the focused middleware components, the integration and improvement of scientific applications to the DCI is also important. Distributed storage management, intelligent workload management, and the analysis framework would be AS' foremost candidates.

4.22. PARTNER: KISTI

AMGA Technology Provider: AMGA, AMGA Manager

KISTI is working closely with KEK Japan for the Belle II experiment to adopt AMGA for its metadata service. Data taking is expected to start in 2015 for 10 years operation. There is also AMGA provisioning discussion with XCEDE. For these communities, KISTI will continue to support AMGA and its related services. There is institutional money to support this activity.

Service stability during peak loads, improve AMGA Python API and federation of AMGA services for redundant master-slave setup are some of the current developments in the pipeline.

5. REFERENCES

R1	Open Grid Service Architecture Use Cases, GFD29, http://www.ogf.org/documents/GFD.29.pdf
R2	Open Grid Services Architecture, Glossary of Terms, GFD44, http://www.ogf.org/documents/GFD.44.pdf
R3	The Open Grid Services Architecture Version 1.5, GFD80, http://www.ogf.org/documents/GFD.80.pdf
R4	UMD Roadmap, https://documents.egi.eu/document/272
R5	EMI Security Architecture, http://openaire.cern.ch/record/5959, http://dx.doi.org/10.5281/ZENODO.5959
R6	Open Science Grid, https://www.opensciencegrid.org
R7	XSEDE, https://www.xsede.org/
R8	DJRA1.2.4, Data Area report, https://cds.cern.ch/record/1277618?ln=en
R9	Cloud Data Management Interface, http://www.snia.org/cdmi
R10	LSDMA, http://www.helmholtz-lsdma.de
R11	GOCDB, https://wiki.egi.eu/wiki/GOCDB/Documentation_Index