**EGI-InSPIRE**

Interim report on the mini projects

**EU MILESTONE: MS801**

|  |  |
| --- | --- |
| Document identifier: | EGI-InSPIRE-MS801-1965-v1.docx |
| Date: | 09/10/2013 |
| Activity: | **SA4** |
| Lead Partner: | **EGI.eu** |
| Document Status: | **FINAL** |
| Dissemination Level: | **PUBLIC** |
| Document Link: | https://documents.egi.eu/document/1965 |

|  |
| --- |
| AbstractThe EGI-InSPIRE SA4 work package has been set up as part of an amendment to the project’s DoW for PY4. This document provides a half-time report on the status of the individual mini-projects for those that last for 12 months, and a final status review for those that have finished after 6 months. |

1. Copyright notice

Copyright © Members of the EGI-InSPIRE Collaboration, 2010. See www.egi.eu for details of the EGI-InSPIRE project and the collaboration. EGI-InSPIRE (“European Grid Initiative: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe”) is a project co-funded by the European Commission as an Integrated Infrastructure Initiative within the 7th Framework Programme. EGI-InSPIRE began in May 2010 and will run for 4 years. This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, and USA. The work must be attributed by attaching the following reference to the copied elements: “Copyright © Members of the EGI-InSPIRE Collaboration, 2010. See www.egi.eu for details of the EGI-InSPIRE project and the collaboration”. Using this document in a way and/or for purposes not foreseen in the license, 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.

1. Delivery Slip

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Name** | **Partner/Activity** | **Date** |
| **From** | Michel Drescher | EGI.eu/SA4 |  |
| **Reviewed by** | **Moderator:** **Reviewers:** <<To be completed by project office on submission to AMB/PMB>> |  |  |
| **Approved by** | **AMB & PMB**<<To be completed by project office on submission to EC>> |  |  |

1. Document Log

|  |  |  |  |
| --- | --- | --- | --- |
| **Issue** | **Date** | **Comment** | **Author/Partner** |
| 1 | 1 Oct 2013 | ToC & Skeleton | Michel Drescher/EGI.eu |
| 2 | 4 Oct 2013 | About half of the mini project contributions added | Michel Drescher/EGI.eu |
| 3 | 9 Oct 2013 | All but one contribution added, Conclusions finished | Michel Drescher/EGI.eu |

1. Application area

This document is a formal deliverable for the European Commission, applicable to all members of the EGI-InSPIRE project, beneficiaries and Joint Research Unit members, as well as its collaborating projects.

1. Document amendment procedure

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE “Document Management Procedure” will be followed:
<https://wiki.egi.eu/wiki/Procedures>

1. Terminology

A complete project glossary is provided at the following page: <http://www.egi.eu/about/glossary/>.

<<The authors should check if the acronyms are covered by the glossary page and if the definition is still correct; all the amendments should be communicated to glossary@egi.eu>>

1. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed − both for coordinating the infrastructure and for delivering integrated services that cross national borders.

The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting ‘grids’ of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example within the ESFRI projects. Additional support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today’s production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

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

The production infrastructure supports Virtual Research Communities (VRCs) − structured international user communities − that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

1. EXECUTIVE SUMMARY

<< The text should provide a summary of the full report so that the reader can ‘in a page’ understand the problem it has been written to cover. This includes an overview of the background material and motivation for the report, a summary of the analysis, and the report’s main conclusions.>>

<<Michel Drescher>>

<<Once all mini project reports are in>>

TABLE OF CONTENTS

1 Introduction 7

2 Mini projects status reports 8

2.1 Work Package management 8

2.2 TSA4.2: Massive Open Online Course Development 8

2.3 TSA4.3: Evaluation of Liferay modules 9

2.4 TSA4.4: Providing OCCI support for arbitrary CMF 10

2.5 TSA4.5: CDMI Support in Cloud Management Frameworks 12

2.6 TSA4.6: Dynamic Deployments for OCCI Compliant Clouds 12

2.7 TSA4.7: Automatic Deployment and Execution of Applications using Cloud Services. 13

2.8 TSA4.8: Transforming Scientific Research Platforms to Exploit Cloud Capacity 14

2.9 TSA4.9: VO Administration and operations PORtal (VAPOR) 15

2.10 TSA4.10: A new approach to Computing Availability and Reliability Reports 17

2.11 TSA4.11: GOCDB Scoping Extensions and Management Interface 19

2.12 TSA4.12: Tools for automating applying for and allocating federated resources 19

3 Conclusion 21

4 References 22

# Introduction

During autumn 2012 the EGI-InSPIRE Project Office identified a number of partners that were under-spending. The EGI-InSPIRE Project Management Board decided to reallocate these unused funds to support supplemental activities that accelerate EGI's strategic goals [R 1] around Community & Coordination, Operational Infrastructure and establishing Virtual Research Environments.

On 14 December 2012 the EGI Project office announced a call for funded mini-projects within the scope and funding regulations of the EGI-InSPIRE project[[1]](#footnote-1). A total of 29 mini-projects were submitted, and by the end of January 2013, the PMB prioritised these and started negotiation with the submitters. In total, 11 mini-projects were funded, while two proposals (“Implementation and testing of central banning in the European Grid Infrastructure”, and “OpenAIRE-based Scientific Publication Repository”) were integrated into existing activities without additional funding.

The funded mini projects are organised and set up as tasks within work Package 8 (SA4) as part of the EGI-InSPIRE project. While regular contributions to the EGI-InSPIRE quarterly reports (the first contribution was made to Quarterly Report 13 [R 3]) focus on summarising the progress made and issues faced in the mini projects, this report serves as a mid-term checkpoint to review the progress so far and compare it to the goals and objectives that were agreed upon during the mini project negotiations. It serves as an opportunity for the mini project leaders and appointed shepherds to reflect on the general mechanics of how to embed the mini projects into the respective context, and adjust how the mini project generally conduct their business.

Section 2 forms the core part of this document. Starting with a brief overview of the management structure of the work package, this section provides status reports of each mini project against its own work-plan and objectives.

Section 3 concludes this milestone document with summarising the overall status of the work package.

# Mini projects status reports

## Work Package management

The Work Package management is split along project and technical management aspects: Four shepherds managing the EGI platforms described in the EGI Platform Roadmap [R 2] take care of providing sufficient context and guidance to the mini-projects so that outputs may be integrated into the EGI production infrastructure as seamless as possible. In practical terms, formal Work Package management is kept at a minimum presuming that mini-projects mostly organise themselves.

Mini-projects utilise the following EGI infrastructure:

* An overview of the mini-projects is maintained in the EGI Wiki[[2]](#footnote-2).
* Mini-projects record their meetings in EGI Indico[[3]](#footnote-3) unless folded into other EGI-InSPIRE meetings.
* DocDB, including an appropriate topic, will be used for permanent documentation[[4]](#footnote-4).
* Weekly reports are collected by WP8 management and relayed to the Activity Management Board, including an executive summary. Through this mechanism, mini-projects are encouraged (and already did so) to report delays, raise issues that require support outside the individual mini-projects, and more. Mini-projects may choose their own reporting frequency (e.g. weekly, bi-weekly) but are required to consistently follow their choice.
* A spreadsheet maps members to mini-projects, and shepherds to mini-projects. It is managed online, and anyone with the link may view it. EGI-InSPIRE PO, shepherds and WP8 management may edit it.

Two mailing lists are provided for mini-project leaders, their deputies and shepherds, and for all mini-project members, respectively, although these are rarely used except for regular weekly report collection. This is not an issue, since all mini-projects are well embedded in their target platform ecosystem.

## TSA4.2: Massive Open Online Course Development

This task concerns the creation of a Massive Open Online Course in which participants will learn to use Grid Computing for their own projects. It focuses on users without any previous large scale computing experience and shows them different methods to use large scale computing facilities. This includes working on a local cluster, using the Grid through the gLite middleware, pilot jobs and workflow management systems.

This project started by creating an outline of the different topics that needed to be taught and determining how the information could be conveyed best. To make the material livelier it was decided to ask people from the EGI community to contribute use cases and show with real-life examples how Grid Computing has facilitated scientific discoveries.

Most of the time for this project has been spent on content creation: creating the slides, examples, movies and animations that will be shown to course participants. Other tasks include the creation of a virtual machine image with which the participants can perform the assignments, the creation of a framework to animate different cluster and grid scheduling methods.

The course will be held from the 18th of November 2013 and lasts six weeks. We will be using the MOOC platform developed for the University of Amsterdam (UvA). After the initial course is over the course material will remain available to the community.

The coming three months we will focus on recording all the lectures, integrating them with the animations, and putting them online. We will also have recording sessions with a number of Dutch scientists who are active in the EGI community. January 2014 will be used to evaluate the course, after which our findings will be presented to EGI.

## TSA4.3: Evaluation of Liferay modules

The objective of the mini-project is to evaluate the Liferay portal with its recently released modules Liferay Sync and Liferay Social Office as a replacement for some of the EGI back office services provided now by CESNET using a set of specialised software systems, and as a web portal platform for the EGI community. The outcome is expected to be best practices and recommendations for the EGI community.

The mini-project is divided among three partners, CESNET which runs the current back office and thus evaluates the service replacement and general portal options, and INFN and SZTAKI, which evaluate compatibility with their community portlets.

### Achieved during the first six months

The INFN partner completed their planned work during the first six months. INFN installed and evaluated two versions of Liferay Social Office 1.5 and 2.0, and the latest version of the Liferay Sync module. They tested interoperability with AAI solutions, namely SAML-based identity federations. They also tested interoperability with the IGI portlets from the community. They also have evaluated Liferay as an alternative for the AppDB (EGI Application Database).

The CESNET partner has assumed most of the tasks in the miniproject. They have installed their own instance of Liferay with Liferay Social Office and Liferay Sync modules, and evaluated the following features:

1. Solution for VRC, VO, NGI, project websites
2. Interoperability with EGI SSO
3. Interoperability and alternative to EGI Helpdesk
4. Interoperability and alternative for Indico
5. Interoperability and alternative for Wiki
6. Interoperability and alternative for DocDB
7. Interoperability and alternative for EGI Blog

The findings were demonstrated on the EGI Technical Forum 2013 conference held in September 2013 in Madrid. However the planned work for this partner is not finished yet and will be continued during the next six months.

The SZTAKI partner has planned one task, evaluation of interoperability of Liferay with Social Office and Sync modules with their SCI-BUS and SHIWA portlets. The have installed their own instance of Liferay with the Social Office and Sync modules, and examined the needed modifications to use the Social Office module with their SCI-BUS and SHIWA portlets.

### To be done in the next six months

The INFN partner has finished their planned work already.

The CESNET partner has evaluated all the planned features on the basic level, but more thorough evaluation is planned in the areas of

1. Implementation of workflows for the Liferay-based alternative for EGI Helpdesk
2. Implementation of so-called *hooks* in Liferay for reimplementation of features of the EGI blog that were required by the EGI and that are present in the current implementation of the EGI blog but that are not provided by the Liferay blog implementation
3. Reimplementation of the current web design of the EGI web site using the tools provided by Liferay, i.e. creating a Liferay theme more close to the original EGI web design

The SZTAKI partner plans to finish their evaluation of interoperability of their SCI-BUS and SHIWA portlets with the Liferay Social Office module.

## TSA4.4: Providing OCCI support for arbitrary CMF

This EGI-InSPIRE mini-project aims to provide a cloud interoperability framework based on OCCI with support for arbitrary cloud management frameworks. One of its key enabling scenarios is to be able to run a predefined virtual machine image at multiple sites of a federated cloud environment and the associated ability to manage the resulting virtual machine. As different cloud management frameworks currently exist and are actively used at different sites, enforcing a particular framework across all sites is neither practical nor desired for a plethora of organizational and technical reasons. Therefore a standardized, uniform interface for the management of virtual machines is needed. The mini-project focuses on the rOCCI framework and rOCCI-server, OCCI implementations used within the EGI Federated Cloud Task, modifying and extending them to improve interoperability with other OCCI implementations in mind.

The mini-project efforts are divided into three main categories:

1. Organization
2. Design and implementation
3. Testing and documented deployment

Tasks in the first category are continuously performed during the whole life-cycle of the mini-project and include status reporting, weekly meetings, team organization and workload management. Tasks in the second category include analysis of the existing code base and other OCCI implementations, identification of required changes, the design of the new rOCCI-server and its implementation in Ruby programming language. The third category also contains continuous tasks including, but not limited to, unit testing, functional testing, integration testing and extensive documentation for all parts of the rOCCI framework and rOCCI-server.

### Milestones for the past six months

**Task 1: Mini-Project Management**

* The team members proposed and agreed on a work schedule, meeting schedule, reporting schedule and development tools, presented the mini-project at EGI Community Forum 2013 and later at EGI Technical Forum 2013.
* All goals in this task have been accomplished.

**Task 2: rOCCI Framework Changes**

* The team members identified changes required to support wider variety of OCCI implementations, primarily OCCI-OS; implemented said changes and deployed updated version of the rOCCI client within the EGI Federated Cloud Task environment.
* As a preparation for the rOCCI-server re-design, the rOCCI framework has been split into three easily maintainable components: rOCCI-core, rOCCI-api and rOCCI-cli.
* All goals in this task have been accomplished.

**Task 3: rOCCI-server Re-design**

* The team members proposed and agreed on a design and started working on implementation of the proposed design in Ruby programming language.
* Goals in this task have only partially completed.

The mini project’s wiki space[[5]](#footnote-5) contains more details.

### Milestones for the upcoming six months

As with the milestones for the previous six months, details are available on the mini-project’s wiki space.

**Task 3: rOCCI-server Re-design**

* Intended completion of goals from Task 3.

**Task 4: Back-ends for CMFs**

* Planning to implement server back-end for OpenNebula and assist with development of the StratusLab back-end.

**Task 5: Testing and Deployment**

* Planning to deploy the new rOCCI-server in a testing environment and later in a production environment of the EGI Federated Cloud Task.

**Task 6: Documentation**

* Planning to provide detailed documentation for deployment, usage and development of the rOCCI framework and rOCCI-server.

### Organizational changes

The original mini-project proposal included unfunded participation of two GWDG members and one former GWDG member. Unfortunately, both GWDG members did not participate at all and the former GWDG member announced his unavailability for this mini-project in M6-M12. This had an impact on the original work schedule agreed on in the beginning of the mini-project. To accommodate this change, we propose the following changes to the original project work plan:

1. Provide support for existing server-side implementations of OCCI via the rOCCI client instead of new back-ends for the rOCCI-server
2. Simplify rOCCI-server architecture by limiting the extent of the back-end abstraction
3. Focus on implementing rOCCI-server back-end for OpenNebula; while assisting with third-party development of the StratusLab back-end
4. Let go of the original intent to implement a native proof-of-concept OCCI client for Java

### Summary

Despite the unexpected organizational changes, this mini-project completed its M1-M6 term without major delays or problems and completed initially stated goals with expected outputs.

## TSA4.5: CDMI Support in Cloud Management Frameworks

This task’s objective is to design and implement a SNIA/ISO CDMI-compliant storage service that integrates with EGI security infrastructure and extends EGI service portfolio by offering an object storage component.

The development aims at offering richer server-side processing functionality to simplify client creation. The initial preparation of this task consisted in setting up a development infrastructure for the project (Github projects[[6]](#footnote-6), RTD documentation[[7]](#footnote-7), CI, Jira). The work plan of the task included the following milestones for the first 6 months:

* **Analysis of user requirements**. Based on discussion on the mailing lists and potential users, we have conclude that the initial plan to support also block devices does not provide much of additional value as this part is anyway performed by infrastructure provisioning services. As such, we have concentrated on the object store aspect. It should be noted that so far we have a problem related to getting real applications to use CDMI solution. We plan to address this once a more stable and feature rich version is done.
* **First prototype version.** A prototype (aka Stoxy) supporting creation and management of object and containers was created, exposing CDMI and explorative ssh interface. The prototype was presented at SNIA Storage Developer’s Conference, as well as co-located CDMI plugfest. To accompany server, also python SDK was created to simplify client integration.
* **EGI AAI prototype integration.** Based on discussion in FedCloud TF mailing list, it was decided to use OpenStack Keystone + VOMS extensions as a way for integrating with the EGI infrastructure. For that, a cryptotoken authentication support was added to Stoxy.

The next six months of the project will focus on these milestones:

* **Community Engagement**. We need to attract actual use cases for a successful continuation of the activity. There are ongoing discussions with several prospective use cases. We plan to create demonstrators for easier client aggregation.
* **Web Client + SDKs**. For better demonstration of capabilities, we plan to create a web client interacting directly with CDMI server. In addition, final versions of Python and Java SDKs with command line tools will be released for user convenience.
* **Final version of Stoxy.** A final version of the server and SDKs are planned to be released at the end of this year with 3 months of the mini-project remaining for testing, bugfixes and validation. The

## TSA4.6: Dynamic Deployments for OCCI Compliant Clouds

<<Marc-Elian Begin>>

## TSA4.7: Automatic Deployment and Execution of Applications using Cloud Services.

This task objective is to design and implement a contextualization capability that supports scientific communities in executing their computing workload through automating the deployment of scientific software on virtual machines using the interfaces and standards used in EGI’s Cloud Infrastructure Platform..

This new capability will allow VRC managers (or advanced users) to define a set of applications that the researchers can easily deploy in virtual machines relieving them from the overhead of setting up the computing environment.

The initial preparation of this task consisted in setting up a new GitHub Project[[8]](#footnote-8) where the members can create repositories for all the project artefacts. The work plan of the task included the following milestones for the first 6 months:

* **Analysis of user requirements and the EGI Federated Cloud testbed**. The team members performed an initial analysis of the support of the EGI Federated Cloud infrastructure[[9]](#footnote-9) and proposed a new extension for the OCCI API to enable contextualization by allowing passing user-provided data to the virtual machines on instantiation[[10]](#footnote-10). In close collaboration with the Particle Physics Phenomenology at CSIC an initial architecture of the service was defined and documented at the GitHub project[[11]](#footnote-11).
* **Initial Implementation of VM contextualization service.** As defined in the architecture, the VM contextualization service allows VRC managers to define applications and the recipes that deploy those applications on the virtual machines; and allows users to query those applications and get the relevant contextualization data for deploying those applications. A first functional prototype of the service is available[[12]](#footnote-12) for testing. This initial version includes support for defining the applications, the recipes and for getting cloud-init[[13]](#footnote-13) compatible contextualization data for the users. The service uses VOMS proxies for authorization as the rest of EGI Federated Cloud Infrastructure.
* **REST API to the service**. A REST interface for the service was defined[[14]](#footnote-14) and implemented for the prototype service. As a proof of concept, the Particle Physics Phenomenology contextualization extension used at CSIC for the OpenStack DashBoard has been refactored[[15]](#footnote-15) to use this API instead of a static list of applications.

The next six months of the project will focus on these milestones:

* **Community Engagement**. With the initial prototype and API available, new user communities will be approached to use the service and gather new requirements. The team members are now in contact with users of the computational chemistry and with a community using Observium[[16]](#footnote-16) for research on monitoring tools.
* **Web Interface**. The API also allows the development of a web interface to the service that will lower the entry-barrier for new users.
* **Integration of Automatic Configuration Tools.** The recipes will be extended to support at automatic configuration tools. Initial tests with Puppet have been already performed.
* **Final version of service.** With the input provided with the new communities, a final version that fixes any issues will be made available to the broader EGI community.

## TSA4.8: Transforming Scientific Research Platforms to Exploit Cloud Capacity

### Goals

The goal of this activity is the derivation of patterns and recipes that can be applied in order to make applications cloud ready. This is done by optimising several use cases that we see most promising to benefit from these actions. The lessons learnt will result in a collection of best practices of which new applications can make use to ease their uptake of cloud technologies. We do this by evaluating existing VM images provided by various user communities and trying to optimize how they make use of cloud resources. Our decisions are supported by questionnaires about the applications targeted at the individual use cases.

### Evaluation phase

We started by doing a more thorough evaluation of the existing images to determine the following:

* What software packages are installed inside the image?
* Is any application specific data shipped with the image and how much?
* Are there any remnants of data that do not need to be shipped with the image, e.g. swap space?

In parallel with the evaluation of the images, we sent a questionnaire to the user communities, asking about the requirements of their application. The topics that we were particularly interested in were: image preparation, workload management, AAI and contextualization, and data handling.

After evaluation of the images and feedback from user communities, we decided on which applications we would focus during the following months. The initial two communities were BioVel and WeNMR. Details about the decisions and proposed actions can be found in our evaluation document [R 4] in the EGI document database.

### Implementation

The key finding during the evaluation phase was that several applications shipped large amounts of data along with the virtual machine images. The order of magnitude observed was several GB. However, communities told us that they only shipped a small portion of their full data sets along with the image for testing. These data sets will be served via cloud object storage or virtual block devices in the future. However, application providers will take a while to adapt to this proposal.

Another deficiency found during the evaluation phase lead us to believe that it is difficult for user communities to craft minimal base images. This is even more difficult when considering the fact that these images are supposed to be run in multiple, heterogeneous resource providers running various virtualization platforms. We have thus started to create basic images that run in several resource providers running different virtualization platforms (KVM/Xen). The creation of such images has been documented. It is planned to publish these basic images in the future, such that user communities have a starting point to base their application specific images on. One of the key aspects of these basic images is contextualization. We proposed to use cloud-init, a widely used framework for initializing VMs in the cloud. It is capable of abstracting away the actual source of contextualization information, which is different in each cloud management framework, behind the concept of a data source.

User communities other than WeNMR and BioVel were not very responsive. However, they have recently re-joined our discussions and we will support them in trying to run instances on EGI federated cloud resources, optimizing their applications according to our findings.

### Next six months

During the next six months, we will provide basic images of popular Linux distributions, equipped with a basic configuration of cloud-init ready to be used by user-communities. The creation of these images will be documented, such that it can be used by other to start from scratch, as well as serve as a starting point for the creation of images for other operating systems.

Regarding data management, we will further advocate the use of cloud storage external to the images, avoiding large payloads of application data.

The entire activity will culminate in a document about patterns and best practices that we have implemented in cooperation with user communities and that can be generally applied to other applications, too.

## TSA4.9: VO Administration and operations PORtal (VAPOR)

VAPOR intends to help small to medium-size grid user communities perform daily administrative and operational tasks, by developing a generic tool to assist community managers and support teams in performing their daily activities. Such communities may typically have no or few dedicated IT support, have scattered scientific activities or fragmented user groups, and may possibly (although not necessarily) make an opportunistic usage of the resources.

The portal is expected to (i) help communities sustain their model by mutualising the administrative and operational cost with other communities, (ii) facilitates the outreach of new user communities by making it easier to start with the administration and operations of a VO.

**Achieved during the first six months:**

Functional specifications:

The starting period (M1 to M3) consisted in a set of phone conferences with partner VOs, in order to define the functional specification of the project, assess existing material that the project may leverage, and sort out priorities in terms of developments. Minutes are available at: [COMPCHEM](https://indico.egi.eu/indico/conferenceDisplay.py?confId=1645), [WeNMR](https://indico.egi.eu/indico/conferenceDisplay.py?confId=1660), [France Grille VO](https://indico.egi.eu/indico/conferenceDisplay.py?confId=1665). This phase resulted in *Deliverable D1 - VAPOR Functional Specifications*[[17]](#footnote-17). Deliverable D1 comes with a companion document that gives development priorities[[18]](#footnote-18): those were sorted by importance for each partner, but also by maturity of the reflection and optionally existing approaches.

Developments:

In a second period (M3 to M6), technical contributions started with the setting up of a development platform at I3S, a [source repository](https://redmine.i3s.unice.fr/svn/vapor/) and [project tracker](https://redmine.i3s.unice.fr/projects/vapor). Technical choices were agreed with the EGI Operations Portal team during a [two-days face to face meeting](https://indico.egi.eu/indico/conferenceDisplay.py?confId=1721).

Developments started as to the features referred to as the *VO Operations management for VO support teams*: resource status indicators and reports, white list of computing elements, report of the list, status and capabilities of all resources supporting a VO by consolidating information from the GOCDB and BDII. This set of features will be completed shortly and is referred to as D3.1 in the figure below.

A last major task initiated in this period is the study and assessment of possible technical solutions to implement the VO Data Management features, referred to as D3.2 in the figure below.

Revision of the project schedule:

The definition of priorities (described above in the starting phase) changed the order of development initially proposed in the project description. As a consequence, D2 (community users management) is postponed after D3 (operations management). D4 (accounting) is deemed less useful and is postponed at the end of the project, if time remains. During the development, D3 appeared to be a bigger work than expected, and it is split into two deliverables: D3.1 (Resource status indicators, statistical reports and white list) and D3.2 (VO Data Management).

The updated schedule is provided in the figure below.



**To be done in the next six months:**

In this period, the first major step will be the development of the VO Data Management (D3.2) features which technical definition has been started in the previous period. This will involve partners CNRS Creatis and GRyCAP, as well as site administrators who showed interest in helping refine the data management procedures.

Then, the last major software package will be the community users management. This task is a quite ambitious feature, in particular because it does not leverage much existing software. As a result, it is unsure, as of today, if this task will realised entirely.

During the last months of the projects, the priority will be put on the deployment of a production-class application properly integrated into the EGI Operations Portal, rather than the development of the Accounting features (D4) that are not deemed essential. D4 is therefore postponed to the end of the project, and will be considered if time remains.

A first deployment phase is planned M8 and will involve the EGI Operations Portal development team. The objective is to make VAPOR accessible to the support team of the biomed VO, so that feed-back can be collected and taken into account.

As of now, the VAPOR software is available under the Apache Licence v2, from the project [SVN repository](https://redmine.i3s.unice.fr/svn/vapor/).

## TSA4.10: A new approach to Computing Availability and Reliability Reports

### Technical Work Plan

The technical work plan includes 4 phases: (1) Requirements assessment, (2) Implementation, (3) Pilot phase and (4) Service deployment.

Initially these phases were scheduled to be sequential. Very early in the project we realized that expecting a sequential project implementation model would be unrealistic as the “Requirements Assessment” process was quite complex. As a result, we opted for a more flexible plan that would allow us to manage the introduced complexity of having the “Requirements Assessment” as an external dependency:



### Requirements assessment

In the new model, although we still retain the 4 projects phases that were already mentioned, we have chosen to split the requirements assessment phase into 3 sub-phases:

* The initial sub-phase was designed as internal to the mini project, given the expertise of the partners with the SAM framework. It was used for kick-starting the mini project for its first six months.
* As planned, the second requirements assessment phase started while implementation had already begun; the EGI Requirements Gathering Task Force gave valuable input during three meetings conducted in July/August 2013[[19]](#footnote-19).
* The and final round of requirements assessing will take place in October/November 2013 and its goal will be to provide input for the reporting capabilities of our product.

### Implementation

Implementation started in May 2013 based on the first requirements assessment captured in the Scrum backlog[[20]](#footnote-20) (the mini project has adopted the Scrum agile project management methodology).

The overall product is split in 4 distinct subsystems: Sync Services, Compute Engine, Web API and WebUI. Currently we are in the middle of the “Implementation” phase and the following set of high level features have already been implemented:

* [Sync Services] Retrieve POEM profiles from POEM Service
* [Sync Services] Retrieve monitoring data from the Brokers
* [Sync Services] Retrieve topology information from GOCDB
* [Sync Services] Retrieve downtime information from GOCDB
* [Sync Services] Prefilter raw monitoring data
* [Compute Engine] Compute status for Service Endpoints
* [Compute Engine] Compute status and A/R for Service Flavors
* [Compute Engine] Calculate A/R for Sites & NGIs
* [Compute Engine] Calculate A/R for NGI Core Services & VOs on Lavoisier
* [Web API] Provide A/R API for integration with Lavoisier
* [Web UI] Distribute A/R results through Lavoisier

Roughly half of the features for the synch services and the compute engine are implemented, while most of the features for the WebAPI and the WebUI subsystems will be implemented in the second half of the mini project. Although this might look different, the project is on plan, as further development on the Web API and Web UI require a sufficiently stable and accurate data set. This is one of the goals of the pilot phase.

### Pilot phase

The pilot phase started on August 2013 with deploying the test bed on GRNET’s ~okeanos cloud platform. One purpose of the “Pilot” phase is to not only providing the testbed but also thorough testing and validating newly developed features, and serve as a demonstration service. Utilising an external, reliable test bed also requires formalised and automated package building processes. This is accomplished using Koji[[21]](#footnote-21).

The second purpose of the pilot phase is to validate the A/R results against the reference data coming from the production service; root cause analysis will serve as further input into the parallel implementation phase. For each Resource Centre in EGI, the absolute differences between this project’s Availability and Reliability figures will be calculated individually, and then compared to the figures coming from ACE – for every month until the deviations are either resolved or within an acceptable range. Currently, validation is underway for 308 resource centres in EGI for the months August and September 2013:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metric | # sites | ε < 1% | 1% < ε < 5% | 5% < ε < 10% | ε > 10% |
| Availability | 308 | 277 | 23 | 7 | 1 |
| Reliability | 308 | 277 | 23 | 8 | 0 |

Accepting deviations of 1% or less, data is already accurate for 90% of all sites for both availability and reliability. On the other hand, data for 97% of the sites is already within the 5% deviation threshold compared to ACE. For the one site where availability differs by more than 10%, the investigation concluded that the raw data available on the two computation engines do not match.

### Production phase

This is the final phase of the mini project. All core features of our product will have been implemented and the deployment of the production infrastructure will start. It is planned to start on February 2014.

## TSA4.11: GOCDB Scoping Extensions and Management Interface

**Objectives Overview**

This project is now complete. The project spanned 6 months starting in April 2013 and finishing in October 2013. This funded a new developer to work with David Meredith to implement the main project deliverables listed below. Both deliverables were completed on time and were integrated into the GOCDB v5 source code. GOCDBv5 was released into production 2nd October.

1. Extend the current ‘EGI’ and ‘Local’ data scoping logic to introduce multiple, non-exclusive scope tags. This allows resources to be grouped into one or more flexible categories such as ‘EGI’ ‘Local’ ‘EGI\_TEST’ and ‘CLIP’.
2. Provide a supporting GOCDB management interface to simplify and speed up daily operational/admin tasks.

The main project task list: <https://wiki.egi.eu/wiki/VT_GOCDBExt>

**Review**

For the most part, the work-plan was followed closely with little deviation. All the main tasks listed at the link above were completed. There is still some documentation to finish but this will be completed over the course of the next few weeks. The project incurred a small overspend.

The end of project review document detailing progress and lessons learnt is at the following link and will not be repeated here: <https://documents.egi.eu/document/1957>

## TSA4.12: Tools for automating applying for and allocating federated resources

This mini project directly supports one of EGI’s key strategic activities, by providing a tool that will allow the provisioning of federated EGI resources for scientific use cases. The tool is built with close collaboration with Resource Allocation Task Force (RATF). The planned phases of the project are presented on Fig. 1.



Fig. 1 Planned phases of the miniproject.

The design phase of the solution in its first version has been finished according to plan in May 2013 and is documented within the RATF wiki[[22]](#footnote-22). Important part of the approved solution is the *pool* concept. Pools are specific declarations of resources providers (NGIs and Sites) that specific allocation scenario might be applicable to their resources. The tool is designed to support three allocation scenarios: free hand to EGI, right-to-revoke and full negation. After a phase of technology assessment the team decided to use the Agreemount framework[[23]](#footnote-23), that supports easy integration, high level of customisation and standardised views for presenting and operating on SLA changes.

The development of version 1.0 started in time and currently is still in progress until end of October 2013. Its current state was presented with a pre-recorded demonstration movie at the EGI TF 2013[[24]](#footnote-24). At the time of writing, the implementation was nearly finished. The full allocation process has been implemented including:

* Resource pool management for resources providers
* Sending resource requests, and
* Automatic resource pool matching.

All negotiation scenarios are implemented and automatic changes are made to OLAs based on the pool description). Support for OLAs allows underpinning them to SLAs. Functions implemented for the Broker include:

1. Dashboard with related actions
2. Visualisation of the SLA status
3. Communication with customer (VO or VO group representative).

Finally an initial set of metrics describing resources was introduced. The remaining functionalities planned for version 1 are the following:

1. Authentication and authorisation integration based on GOCDB (for providers) and VO id cards (for VO),
2. Pools usage management as well as integration with EGI Operation Portal.

All the works listed above are currently under development. By end of October the Resource Providers and VOs participating in the RATF will test the developed tool.

Parallel to implementation and integration works, the design phase for the final version of the tool was initiated according to plan. The list of the most crucial extensions has been identified (see below) and design activity will focus on them. The first goal is integration with Scientific Review process. At the time the design of the first version was prepared, the Scientific Review process was not yet established nor technically planned. Therefore any necessary interfaces require design and implementation very soon. The second important challenge is to prepare the system for more than one federation layer. Programmable service interfaces for federated providers, which would enable interoperation with other tools may support this requirement. Some further inputs from the RATF need finalisation and the tool implementation need to adapt accordingly (the current implementation was based on a draft version). Additionally, the analysis of new use cases that are under discussion in RATF (e.g. CTA VO signing SLAs with sites already supporting their experiment) would be source of further requirements, and some of them need to be included in the version 2.

The design phase will conclude in November 2013, immediately followed by implementation works. Version 1 will be used for selected scenarios in EGI production environment.

In summary, the mini project is progressing according to the plan and will achieve its goals. The important part of the success is collaboration with RATF, which should finalise the results needed for tool in the appropriate time, and with SA1, where the related operational processes need to be constructed and put into operation. SA1 already performed the action on first pools collection, by sending a pool survey to sites and NGIs. Those pools definition will be transferred to RA tool as start-up set-up.

# Conclusion

All funded mini-projects were organised by re-using as much project administration infrastructure as possible: Work Packages 8 (SA4) was set up to last 12 months with each mini project represented as a task within. EGI SSO, Wiki, PPT2, Indico and DocDB are used for administrative purposes. Although available, mini projects were not constrained to use these tools: any alternative agreed by task members that were openly available at no additional cost (i.e. not claimable on project funding) was allowed to be used.

No particular management or administration structure was required; every mini project managed itself with one identified person acting as the main contact point for coordination with work package administration and shepherd.

This separation of formal management and technical coordination is very similar to the matrix management approach popular in the commercial world. It has proven very successful and overcame a common pitfall in matrix management[[25]](#footnote-25) in that the roles of the work package leader and the shepherds were defined very clearly in the beginning, avoiding for example conflict of loyalty, and increased costs: Work package management is responsible for formal administration of the mini projects and ensuring that contractual obligations (within the EGI-InSPIRE project) are fulfilled in form of milestone document production and contributions to quarterly reports. Shepherds, on the other hand, steer mini projects so that their technical output is of maximum use to EGI as a whole. Shepherds oversee the scope and objectives of a mini project, and embed their activities in the target domain in EGI: Initially, the shepherd role was spread over four individuals; recently this has been aggregated over three shepherds. These represent, broadly speaking, the three EGI platforms described in the EGI Platforms Roadmap [R 2]: EGI Core Infrastructure Platform, EGI Cloud Infrastructure Platform, and EGI Collaboration Platform.

This close collaboration with the target domain was key for the success and maintained scope of all mini projects until the time of writing of this milestone. Connecting the mini projects to EGI.eu’s community and coordination services, the mini projects were able to focus as much as possible on their technical work, active technical participation in EGI forums and other interoperation activities while benefiting from the public relations support provided by EGI.eu.

Considering the current success of the mini projects, this model may serve as a blueprint for programme management in EGI.

# References

|  |  |
| --- | --- |
| R 1 | EGI Strategic Plan, Dx.y, DocDB |
| R 2 | EGI Platform Roadmap, MS514, <https://documents.egi.eu/document/1624>  |
| R 3 | EGI-InSPIRE Quarterly Report 13, MS127, <https://documents.egi.eu/document/1928>  |
| R 4 | Evaluation and work plan for Cloud Capability use cases, <https://documents.egi.eu/public/ShowDocument?docid=1824>  |
| R 5 |  |

1. <https://mailman.egi.eu/mailman/private/inspire-taskleaders/2012-December/000106.html> (might require login) [↑](#footnote-ref-1)
2. <https://wiki.egi.eu/wiki/Overview_of_Funded_Virtual_Team_projects> [↑](#footnote-ref-2)
3. <https://indico.egi.eu/indico/categoryDisplay.py?categId=93> [↑](#footnote-ref-3)
4. <https://documents.egi.eu/public/ListBy?topicid=51> [↑](#footnote-ref-4)
5. <https://wiki.egi.eu/wiki/VT_OCCI_for_CMF#Tasks> [↑](#footnote-ref-5)
6. https://github.com/stoxy [↑](#footnote-ref-6)
7. https://stoxy.readthedocs.org/ [↑](#footnote-ref-7)
8. <https://github.com/AppDeployment> [↑](#footnote-ref-8)
9. <https://github.com/AppDeployment/documents/blob/master/cloudsupport.md> [↑](#footnote-ref-9)
10. [https://wiki.egi.eu/wiki/Fedcloud-tf:WorkGroups:Contextualisation#OCCI\_extension](https://wiki.egi.eu/wiki/Fedcloud-tf%3AWorkGroups%3AContextualisation#OCCI_extension) [↑](#footnote-ref-10)
11. <https://github.com/AppDeployment/documents/blob/master/architecture.md> [↑](#footnote-ref-11)
12. <https://193.146.75.143:5000/> [↑](#footnote-ref-12)
13. <http://cloudinit.readthedocs.org/> [↑](#footnote-ref-13)
14. <https://github.com/AppDeployment/documents/blob/master/api.md> [↑](#footnote-ref-14)
15. <https://github.com/AppDeployment/feynapps> [↑](#footnote-ref-15)
16. <http://www.observium.org/wiki/Main_Page> [↑](#footnote-ref-16)
17. https://wiki.egi.eu/wiki/VT\_VAPOR:VAPOR\_features\_description [↑](#footnote-ref-17)
18. https://wiki.egi.eu/wiki/VT\_VAPOR:VAPOR\_features\_priorities [↑](#footnote-ref-18)
19. https://docs.google.com/document/d/1m2gECk57mygsafXPdNZNtFaUOpHyhPwmG4t1OTsyxVs [↑](#footnote-ref-19)
20. https://code.grnet.gr/projects/ar-ng/issues?query\_id=91 [↑](#footnote-ref-20)
21. https://koji.afroditi.hellasgrid.gr/koji/packages?tagID=120 [↑](#footnote-ref-21)
22. <https://wiki.egi.eu/wiki/Resource_Allocation_Task_Force> [↑](#footnote-ref-22)
23. <http://agreemount.com/>, to be published soon. [↑](#footnote-ref-23)
24. The movie is available on-line at <https://dl.dropboxusercontent.com/u/62958702/EGIBazaar-demo-v1_1.mp4> [↑](#footnote-ref-24)
25. http://en.wikipedia.org/wiki/Matrix\_management#Current\_thinking\_on\_matrix\_management [↑](#footnote-ref-25)