

# EGI-InSPIRE

## HUC SOFTWARE ROADMAP

### EU MILESTONE: MS612

---

|                      |                                                                                           |
|----------------------|-------------------------------------------------------------------------------------------|
| Document identifier: | EGI-MS612-684-VO                                                                          |
| Date:                | <b>05/10/2011</b>                                                                         |
| Activity:            | <b>SA3</b>                                                                                |
| Lead Partner:        | <b>INFN</b>                                                                               |
| Document Status:     | <b>FINAL</b>                                                                              |
| Dissemination Level: | <b>PUBLIC</b>                                                                             |
| Document Link:       | <a href="https://documents.egi.eu/document/684">https://documents.egi.eu/document/684</a> |

---



### Abstract

This document provides the third overview of the Roadmap for the development and deployment in the reference user communities of the software included in EGI-InSPIRE SA3. The Heavy User Communities (HUCs) who are part of the project are the primary target of the document, which is intended to give them information on the features available now and in the future, and offer the opportunity to interact with the planned developments so that they can best fit their needs.

However the document is open to other EGI-InSPIRE users and potential users, who may be interested in adopting parts of the software for their uses and in suggesting developments to this effect.

All the Heavy Users Communities included in the SA3 work-package have contributed to this report on the HUC Software Roadmap with information on the status and planning of their services and tools. New versions of this document are produced every 6 months, starting from MS602 [MS602] in Project Month 4. The communication between the different communities has started and is growing – this is a visible and concrete deliverable of the project where it is clear that the funding model and goals of the project are succeeding in motivating common tools and services, even if at the architectural but sometimes also at the implementation and deployment level; the planning for their software of potentially more general interest, has been exposed to the other communities of heavy users and to the general users communities.

## I. COPYRIGHT NOTICE

Copyright © Members of the EGI-InSPIRE Collaboration, 2010-2014. See [www.egi.eu](http://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](http://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.

## II. DELIVERY SLIP

|             | Name                                                                    | Partner/Activity | Date      |
|-------------|-------------------------------------------------------------------------|------------------|-----------|
| From        | Laura Perini, Jamie Shiers                                              | INFN/SA3         | 16-8-2011 |
| Reviewed by | <b>Moderator:</b> Mathilde Romberg<br><b>Reviewer:</b> Pawel Wolniewicz | DMSU<br>MAN      | 24-8-2011 |
| Approved by | AMB & PMB                                                               |                  | 4-10-2011 |

## III. DOCUMENT LOG

| Issue | Date       | Comment                                | Author/Partner                         |
|-------|------------|----------------------------------------|----------------------------------------|
| 1     | 16/08/2011 | First draft                            | Laura Perini/INFN<br>Jamie Shiers/CERN |
| 2     | 27/09/2011 | Revised Draft after Reviewers comments | Laura Perini/INFN<br>Jamie Shiers/CERN |

## IV. 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.

## V. 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>

## VI. TERMINOLOGY

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

## VII. 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 European Strategy Forum on Research Infrastructures (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.

## VIII. EXECUTIVE SUMMARY

This report provides a snapshot of the status and planning (in the first half of year 2011) of the services and tools developed and supported for the needs of the Heavy User Communities (HUCs): High Energy Physics, Life Sciences, Astronomy and Astrophysics, and Earth Sciences. They have provided information both on the software that is currently specific for each of them and for the software that is already of interest for more than a single community.

This last category consists of the same services and tools considered in the previous versions of this Milestone: the Dashboards, GANGA and related applications, the HYDRA and GrelC services, the Kepler, Gridway, SOMA2 workflow schedulers, and the enabling of MPI applications, which receives important contributions also by the Computational Chemistry HUC.

This report is a checkpoint for the goals of SA3, recalled briefly in the points below.

- To transition the services and tools from the communities that have already adopted DCIs, to where their services are part of the general service infrastructure provided through EGI or are sustained by other means – either within their own community or through external software providers (e.g. middleware projects such as EMI).
- To use the experiences obtained by these early adopting communities in integrating new data sources, tools and services to improve the experience for all user communities.
- To ensure that all the user communities supported by EGI should experience no disruption as they move from their current e-Infrastructure provider.

The report and planning of the services and tools, sketched in this report, show relevant developments and widespread usage, mainly still concentrated in one or two communities. The relevant effort going toward making easier the use of these services and tools, while at the same time increasing their functionalities as required by the users, promises well for their more general use within SA3 and also outside it.

## Table of Contents

|          |                                                                            |           |
|----------|----------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>INTRODUCTION.....</b>                                                   | <b>7</b>  |
| <b>2</b> | <b>ROADMAP FOR THE SA3 SHARED SOFTWARE SERVICES AND TOOLS .....</b>        | <b>8</b>  |
| 2.1      | DASHBOARDS .....                                                           | 8         |
| 2.2      | APPLICATIONS .....                                                         | 9         |
| 2.2.1    | GANGA.....                                                                 | 9         |
| 2.3      | SERVICES .....                                                             | 10        |
| 2.3.1    | GrelC.....                                                                 | 10        |
| 2.3.2    | HYDRA.....                                                                 | 11        |
| 2.4      | WORKFLOW AND SCHEDULERS.....                                               | 12        |
| 2.4.1    | Kepler and GridGridGridWay .....                                           | 12        |
| 2.4.2    | SOMA2 .....                                                                | 13        |
| 2.5      | MPI.....                                                                   | 13        |
| <b>3</b> | <b>ROADMAP FOR THE COMMUNITY SPECIFIC SOFTWARE .....</b>                   | <b>15</b> |
| 3.1      | SERVICES FOR THE HEP VIRTUAL RESEARCH COMMUNITY.....                       | 15        |
| 3.1.1    | HammerCloud .....                                                          | 15        |
| 3.1.2    | CRAB and CRAB Analysis Server .....                                        | 15        |
| 3.1.3    | Data Management .....                                                      | 16        |
| 3.1.4    | Persistency Frame-work.....                                                | 17        |
| 3.1.5    | DIRAC .....                                                                | 18        |
| 3.2      | GRIDGRIDSERVICES FOR THE LIFE SCIENCE VIRTUAL RESEARCH COMMUNITY .....     | 18        |
| 3.3      | SERVICES FOR THE ASTRONOMY & ASTROPHYSICS VIRTUAL RESEARCH COMMUNITY ..... | 19        |
| 3.4      | SERVICES FOR THE EARTH SCIENCE VIRTUAL RESEARCH COMMUNITY .....            | 20        |
| <b>4</b> | <b>CONCLUSIONS.....</b>                                                    | <b>21</b> |
| <b>5</b> | <b>REFERENCES .....</b>                                                    | <b>22</b> |



## 1 INTRODUCTION

This document provides the third overview of the Roadmap<sup>1</sup> for the development and deployment in the reference user communities of the software included in EGI-InSPIRE SA3. The Heavy User Communities (HUC) who are part of the project are the primary target of the document, which is intended to give them information on the features available now and in the future, and offer the opportunity to interact with the planned developments so that they can best fit their needs.

However the document is open to other EGI-InSPIRE users and potential users, who can be interested in adopting parts of the software for their uses and in suggesting developments to this effect.

This document reflects the present status in the elaboration of a full roadmap (the HUC Software Roadmap is a Milestone due periodically in SA3): most of the software developments are still concentrated toward a single community, however much effort is devoted to the documentation and to the easiness of use of the different products, both necessary conditions for attracting new user communities. For the different software products the planning included in this document covers at least the main features foreseen in the next year, in some cases providing well defined internal milestones and initial indications for longer term developments.

---

<sup>1</sup> Previous roadmaps can be found in MS602 [MS602] and MS606 [MS606].

## 2 ROADMAP FOR THE SA3 SHARED SOFTWARE SERVICES AND TOOLS

### 2.1 *Dashboards*

The Experiment Dashboard [DASHBOARD] framework provides monitoring of the WLCG infrastructure from the perspective of the LHC experiments and covers the full range of their computing activities, such as data transfer, job processing and site commissioning. The system has proved to be an important component of the computing infrastructure of the LHC experiments, which rely on the Dashboard applications for monitoring data distribution, data processing and the usability of distributed sites and services. Being initially developed for the LHC experiments, the Experiment Dashboard offers common solutions in the area of monitoring the large scale computing activities on the Grid.

Many of the Experiment Dashboard applications are generic and can be used by other communities, under the condition that their data management and workload management frameworks are instrumented for job status reporting both for data transfer and data processing tasks. For instrumentation of the data management and workload management frameworks, the Experiment Dashboard provides a generic library for reporting monitoring data via Messaging System for the Grid (MSG) based on ActiveMQ message broker.

The Site Status Board application, which is successfully used by ATLAS and CMS Virtual Organizations (VOs) for distributed computing shifts, provides a generic framework for storing and managing monitoring metrics with a flexible user interface, which can be easily customized according to the needs of a particular user community.

Along with the wide range of the monitoring applications, the Experiment Dashboard offers a generic framework useful for other large scale software systems. The framework includes building blocks for the handling of monitoring data flow, which includes data collectors, data storage and visualization. The framework was successfully used for software systems external to the Experiment Dashboard, for example ATLAS Distributed Data Management. The framework has proved to facilitate the software development process and to improve deployment, maintenance and support of the framework-based applications.

During the second half of 2011, the Experiment Dashboard framework will evolve further, in particular with regard to new visualization technologies. New architectural principles aim to fully decouple data storage implementation from client-side AJAX-based user interfaces. This will not only allow for improvements in the software development process but it will also contribute to the sustainability of the Dashboard applications, since changes in the data storage implementation will not require changes in the user interface. Several Dashboard applications are being redesigned following these architectural principles, among them are ATLAS DDM Dashboard and Site Status Board. The first prototypes of the new versions of the ATLAS DDM Dashboard and Site Status Board provide faster, richer and more intuitive user interfaces and they have received very positive feedback from the user communities.

Both applications will be deployed in production by the end of 2011.

The redesign of the SAM system required reimplementing of the Site Usability Interface (SUM), which is used by the LHC VOs in order to evaluate the performance of distributed sites and services



from the VO perspective. During 2011, development of the new SUM application progressed well. In the middle of summer 2011, the first version with complete functionality was deployed to the test server for validation by the user communities. By the end of 2011, it should be deployed in production. SUM application is one of the common monitoring solutions which can be used outside the LHC scope.

## 2.2 Applications

### 2.2.1 GANGA

Ganga [GANGA] is an easy-to-use frontend for job definition and management that provides a uniform interface across multiple distributed computing systems. It is the main end-user distributed analysis tool for the ATLAS and LHCb experiments and provides the foundation layer for services managing large numbers of jobs such as HammerCloud [HAMMERCLOUD]. During the average month Ganga is utilised by around 500 unique users who are dispersed around 60 different research institutes and sites.

CERN leads the Ganga Core development and contributes to ATLAS-specific Ganga functionalities. The Core part of Ganga is a software code-base which is shared between different user communities. The Ganga Core software is driven by the evolution of requirements of Heavy User Communities (HUCs), prevalently of High Energy Physics (HEP), but also of other communities. This is performed in several ways:

1. Missing capabilities and features of general interest are developed in the Core software and are made available to HUCs,
2. Core software is adapted to the needs of HUCs by providing better foundation and integration for HUC-specific features,
3. Those features out of the Core ones, that are of general interest are refactored into the Core software to increase code sharing and consistency of the Core software environment

On-going Core development projects include:

- Migration of the GangaATLAS 'prepare' functionality into the Core, to provide a generic solution for persisting the state of application objects.
- Development of a generic solution for handling the output of data, which gives users more control over where, and in what form, the output from their jobs is stored.
- Improvement of the pre-release software testing mechanism, providing more reliable test results and thus improving the end-user experience.
- Consolidation of project web pages, update of project documentation and deployment of a Ganga-developers blog, providing a forum in which experts hints and tips can be shared.
- Integration of the software into standard Unix distributions
- Refactoring and integration of community-specific Ganga plugins (such as “tasks” for the ATLAS VO) into the common core framework.
- Improvements in the job-submission speed will be investigated, which will be of particular benefit to users who submit large batches of jobs.

Additionally, the ATLAS experiment modules will undergo evolution over the next six months, including:

- Improvements to the MultiTasks package, enabling job chaining and automatic collection and merging of output datasets.
- Enhanced support for DQ2 Tag datasets and files, as generated by the ELSSI user interface.
- The output merging algorithm will be tuned to provide improvements.

In addition to the above, there are significant contributions to Ganga Core that are not funded by EGI-InSPIRE and which include improved support for automatic merging and resubmission and refactoring of ATLAS tasks into a common framework. This contribution are fully coordinated within Ganga and may in future turn out of interest also outside ATLAS.

## 2.3 Services

### 2.3.1 GRelC

The GRelC [GREL] service is a Grid database management service aiming at providing access and management functionalities related to relational and non-relational databases in a Grid environment.

During the EGI-InSPIRE project, the GRelC system will be enhanced to support the EGI communities with a new set of functionalities available as web application both through the GRelC Portal and the DashboardDB interface.

In particular the new releases of the GRelC Portal will help users in easily and transparently port in Grid existing DBs, manage several GRelC service instances, Grid users, data access policies, etc.

Support in terms of management, monitoring and control of the GRelC services will be provided through the DashboardDB. The DashboardDB will provide several views (including charts, reports, tables) about the GRelC deployment, the status of the services, the list of available Grid-databases, the supported VOs, related keywords/tags, user's comments, etc. The design of this web application started during the Y1 and a preliminary version has been tested and released in July 2011. The infrastructure behind the DashboardDB (back-end part) has been designed, implemented and tested during Y1 according to the roadmap, so there have been no deviations with regard to the initial plan. The future plan related to this part of the DashboardDB infrastructure basically focuses on doing test and bug fixing. No major changes are foreseen.

From a web application part point of view, the DashboardDB allows to monitor the GRelC instances deployed in the EGI Grid. During Y2 new GRelC instances will be registered to the DashboardDB to have a complete view of the deployment status. Such a monitoring part will complement the existing monitoring tools already available in the EGI environment providing a different perspective more focused on "Grid-database management" services.

The work on the DashboardDB will continue during the second year. By the end of the second year a second major release (v2.0) of the DashboardDB including the Grid-database registry will be available online. Anyway during Y2 several versions will be incrementally released.

The Grid-database registry will contain all of the information about the Grid-databases available in the GRelC system, the associated VOs, a description, some keywords and other useful metadata. During the third year of the project, the final release (2.2) will basically offer little new functionalities, but a higher level of robustness. The registry will complement the functionalities provided by the EGI Application Database and will represent a distributed and multi-VO system. In brief, the DashboardDB will provide a *cross-VO Grid-database system* giving the user community both a unified view about the available Grid-databases and several tools to interact with them.

The design of the internal modules of the DashboardDB web application performed during Q3 took into account the Web2.0 paradigm. Mashup, Google Maps, permalinks, comments, are just some of the features that were considered during the design phase. During Y2, the implementation will proceed accordingly, implementing "export" (permalink) capabilities and "map-based" (through the adoption of Google Maps) interfaces. Specific charts in the DashboardDB will focus on Grid-database management aspects like the number of registered databases, the supported VO, etc.

An important task carried out during Y1 was the SA3 EGI Questionnaire. The questionnaire aims at providing an up to date list about databases (relational, XML-based, etc.):

- Already in place but that need to be ported in Grid in the context of the EGI-Inspire project.

- Already ported in Grid and so accessible in the context of the EGI-Inspire project.
- Not yet deployed in the context of the EGI-Inspire project.
- Available from external sources via FTP, HTTP and that would need a Grid-enabled instance and interface in the context of the EGI-Inspire project.

A new (extended) version of this questionnaire will be sent to the users during Y2 once the registry will be available online and the users could be attracted more by the new tool. So far (during Y1), few feedback have been collected through the questionnaire, but they were really useful to start implementing new use cases, in particular in the LS context. In this case, as reported in QR5, new biological data sources have been ported in Grid, with the long-term goal (during Y2 and Y3) to set-up a larger set of biological databases for the bioinformatics community.

Concerning the Earth Science community, the GRelC service will continue to be used in a couple of use cases related to the Euro-Mediterranean Center for Climate Change (CMCC). In this context, the GRelC services will be exploited to manage metadata information stored into both relational and XML databases. Specific performance evaluation tests on large XML metadata resources will be carried out during Y2 to evaluate scalability issues and identify optimization strategies. Support will be also provided through the Climate-G portal, the scientific gateway of the Climate-G testbed. During Y2 and Y3, additional use cases will be defined, according to the user's requests and the feedback on the SA3 Questionnaire.

### 2.3.2 HYDRA

Hydra is a file encryption/decryption tool developed at CERN as part of the gLite middleware, enabling encryption of sensitive files stored on storage resources [HYDRA].

An experimental Hydra service has been successfully deployed at the end of Q4 (April 2011), on a gLite release 3.1 UI, and is usable for test purposes. Work is ongoing to migrate Hydra to gLite release 3.2 at CERN by gLite experts from the EMI project. A three servers-based Hydra key-store will be deployed as a service for the life sciences community. This task, due in the first year of the EGI-InSPIRE project, has been delayed as it is deemed preferable, for long-term maintenance issues, to wait for the gLite 3.2 version of Hydra to be released: hence the service shall be deployed on recent gLite 3.2 servers from the beginning, rather than have to move installed key stores on new machines later on.

The schedule has been updated to take into account the delay due the late releasing of the Hydra software for gLite 3.2. The service delivered today is a test Hydra catalogue mostly used for the validation of the functionality delivered, including:

- Encryption key registration and removal.
- Encryption key access control.
- File encryption / decryption from User Interfaces and Worker Nodes.

Three servers dedicated to hosting the service have already been provisioned on 3 different locations in France: CNRS I3S (Sophia Antipolis), CNRS Creatis (Lyon), HealthGrid (Clermont-Ferrand).

For the coming year, the provisional schedule in the production service delivery is:

- October 2011 (PM18): delivery of a distributed Hydra service ready for production use within the Life Science community. (This deadline depends on the availability of the Hydra package for version 3.2).
- Unless specific intervention is required due to changes in the gLite middleware Data Management System, a functionality check is then planned every semester from PM20:
  - December 2011 (PM20): revalidation of the functionality delivered.

- July 2012 (PM27): in preparation for D6.7, revalidation of the functionality delivered with and scale tests (encryption overhead measurements, number of keys that can be stored and number of accesses that the Hydra catalogue can handle).
- January 2013 (PM33): revalidation of the functionality delivered.

## 2.4 Workflow and Schedulers

### 2.4.1 Kepler and GridGridGridWay

As indicated already in MS606, Kepler [KEPLER] is a software application for the analysis and modelling of scientific data. Kepler allows scientists to create their own executable scientific workflows by simply dragging and dropping components onto a workflow creation area and connecting the components to construct a specific data flow, GridWay [GRIDWAY] is a Metascheduler that automatically performs all the **submission steps** and also provides the runtime mechanisms needed for dynamically adapting the application execution.

In the previous period several user workflows have been designed, implemented and supported. Among them, as described already in the annual report, it is worth to mention for example workflow is a combination of the equilibrium solver CHEASE supplying input information to the MHD stability code MARS-F that calculates complex RWM eigenvalue (growth rate and real frequency). This use case represents a kind of parametric+parametric execution model, where each application is executed following a parameter scan approach. Another example is VMEC - DKES (1) case, the workflow also follows kind of parametric+parametric model. The number of different cases explored by the parameter scans can be especially large in the case of the VMEC application, where many different parameters can be studied. DKES measures the transport of particles using an equilibrium previously calculated by VMEC.

During the next period the support for the workflows will be continued. The work and tests of the VMEC-DKES will be continued. Since simplest case, for each of the runs fire more than 2500 jobs (and the cases can only be increased) optimisation of the workflows are under implementation (for example to base workflows on statuses of jobs using LB events level).

The next workflow which is going to be developed is the FAFNER+ISDEP one. This use case will allow studying the movement of particles in plasma. Both applications will run using parametric scans. Once several runs of FAFNER have finished, the results are collected and sent to ISDEP. Again, several runs of ISDEP are launched and the results are collected and combined.

We plan on running the first applications from Kepler using the GridWay metascheduler during the next six months. There are two possibilities to support GridWay:

- 1) Support the command line interface provided by GridWay (with the option of using web services).
- 2) Support the DRMAA API directly from Kepler.

We will start by supporting the command line interface. From the previous use cases implemented, VMEC is already prepared to be launched using GridWay. Also, from the new use case FAFNER+ISDEP, FAFNER is been managed using the GridWay commands. These two applications will offer the possibility to check the new developments.

Among the future use cases, those using GEM will require of the DRMAA, since this application has been developed using the API. Once the support to DRMAA is implemented, those workflows will be arranged.

The activity has been also focused on establishing collaboration with other user groups interested on the actions being carried out by the activity and such approach will be continued in next period. For

example the collaboration with the Chemistry community will be continued in order to establish first workflow examples.

## 2.4.2 SOMA2

SOMA2 [SOMA2] is a versatile modeling environment for computational drug discovery and molecular modeling. SOMA2 is operated through a WWW-browser and it offers an easy access to third-party scientific applications. The SOMA2 environment offers a full scale modeling environment from inputting molecular data to visualization and analysis of the results, and including a possibility to combine different applications into automatically processed application workflows.

During project quarters 4 and 5, CSC has maintained and operated CSC's SOMA2 service. Also, we have started to integrate scientific application Autodock 4 into SOMA2 making use of the Grid support features. In addition, we have investigated how to setup a SOMA2 service which would be provided to other user communities as well and not only for current CSC users. SOMA2 system was also presented to the EGI communities in EGI User Forum 2011 in April. Program development effort has been put in upgrading the used core UI library components and in migration work to be able to use the updated components. Work has been done to improve the basic Grid support in SOMA2 along with other minor fixes and improvements.

For the upcoming six months, we still plan to be able to release a Grid-enabled version of SOMA2. Also, we intend to finish a SOMA2 capsule for the chosen pilot application Autodock 4. This will enable to set up a Grid-enabled application service, initially on a national level, in CSC's SOMA2 provided for Finnish academic researchers and after this we will expand the service so that it will be available for other user communities as well. We will also operate and support the SOMA2 service. In addition, we will advertize the upgraded SOMA2 service to existing user communities.

Our longer term goal is to expand the scientific applications selection in the SOMA2 service, and integrate application services from different Grid entities into SOMA2. This should be set as an important milestone because from the end users point of view, this would make using scientific applications in different Grids very easy and transparent. During the third project year, we will continue to maintain and operate the SOMA2 service and seek possible scientific applications to be added as part of the service. During all project years, we will support using SOMA2 service. Also development of SOMA2 gateway will continue according to feature roadmap including possible feature requests from the user communities bug fixes and other enhancements.

## 2.5 MPI

The MPI [MPI] sub-task shall focus on a number of core objectives over its 36 month period (Y1 - Y3). Many of these objectives are iterative, often requiring updates or fine-tuning. Other objectives, including participation in the EGI Community and Technical forums, shall be repeated at regular intervals. The core sub-task objectives are:

- Improved end-user documentation, addressing MPI application development and job submission in ARC, gLite and UNICORE,
- Quality controlled MPI site deployment documentation,
- Outreach and dissemination at major EGI events and workshops,
- User community, NGI and site engagement, gathering direct input,
- Participation in selected standardisation bodies.

This sub-task has produced numerous MPI workbenches of increasing complexity with specific high impact on the Computational Chemistry and Fusion communities. These products include parallel implementations of Linear Algebra routines (produced by UNIPG). These have direct reuse by many



other communities. UNIPG have also worked on implementing the “horse race” programs for atom diatom quantum reactive scattering in UNIPG ABC and RWAVEPR<sup>2</sup>. This work shall continue over the PY2 period. This code has been successfully implemented on two different GPGPU architectures. In addition, UNIPG shall continue exploration of a parallel implementation of Chimere. Work on parallel implementation of DL\_POLY; NAMD; VENUS96 was completed in Y1, Work on GROMACS and GAMESS is in a testing phase, and will be continued in Y2.

Some improvements to the MPI documentation<sup>3</sup> were also made in Y1. The EGI Wiki now provides the definitive source of information for MPI support. The first release of the UMD gLite-MPI package is expected by Project Month 16, and will require proof-reading and updates to the existing documentation. Furthermore, documentation on the ARC and UNICORE still require attention. It is expected, that the first versions of ARC and UNICORE MPI documentation will be available by PM 18 as more UMD products become available on the production infrastructure. In particular, the release of the UMD glite-WMS and the glite-MPI products are expected to contain significant middleware changes required to ensure the correct support for many MPI job types. For example, these include:

- User defined allocation of processes/nodes.
- Direct Support for OpenMP codes.

This work is under the leadership of CSIC, and is aimed at steering MPI and Parallel jobs technical requirements.

Both UNIPG and TCD have started work on exploiting GP-GPUs. The UNIPG work is already at an advanced stage: An entire job flow that enables the Local Resource Management System (LRMS) to discriminate the GPU resources requests, through Glue Schema parameters, has been defined in order to allocate, in a dynamic fashion, the required resources on a Cloud-like infrastructure. This solution helps overcome some weakness in many job schedulers, which do not currently support GPU resources. TCD is likely to heavily use StratusLab – an open-source cloud distribution that allows Grid and non-Grid resource centres to offer and to exploit an “Infrastructure as a Service” cloud.

As part of User Community engagement effort, the MPI team will regularly survey Virtual Organisations, Users and Site administrators for critical feedback. This will also act as a means to gather information about current deficits and future requirements. The first survey is expected to be prepared in project month 16.

---

<sup>2</sup> See R. Baraglia, M. Bravi, G. Capannini, A. Lagana', E. Zambonini, A Parallel code for time independent quantum reactive scattering on CPU-GPU platform, Lecture Notes Computer Science 6784, 412-427 (2011); and L. Pacifici, D. Nalli, D. Skouteris, A. Lagana', Time dependent quantum reactive scattering on GPU, Lecture Notes Computer Science 6784, 428-441 (2011)

<sup>3</sup> [https://wiki.egi.eu/wiki/MPI\\_User\\_Guide](https://wiki.egi.eu/wiki/MPI_User_Guide)

## 3 ROADMAP FOR THE COMMUNITY SPECIFIC SOFTWARE

### 3.1 *Services for the HEP Virtual Research Community*

Services for High Energy Physics are covered in detail in MS610 [MS610].

In the following sections a brief intermediate report and planning is provided covering the next 6 months, the schedule on which this document is updated and replaced.

#### 3.1.1 HammerCloud

[HammerCloud] is a Grid site testing service developed around Ganga. HammerCloud uses frequent short jobs to validate a site's availability and functionality, and also delivers on-demand stress tests to aid in site commissioning or general benchmarking. Ganga was developed with the ATLAS LHC experiment and that collaboration continues to use the service actively. The CMS and LHCb LHC experiments also have HammerCloud plug-ins developed.

Achievements in 2011 include:

- Complete the deployment of HammerCloud for CMS and LHCb.
- Deployment of the auto-exclusion system for the ATLAS experiment that automatically excludes sites from the distributed analysis systems. Improvements in reliability have been measured, reducing the failure rate of Grid jobs up to 50%.
- Extension of functional testing to production queues for ATLAS.

Plans for 2012 include:

- Migration of the ATLAS instance to the new version, merging development with the CMS and LHCb instance.
- Development of features for new database backends based on NoSQL engines (such as CouchDB) for the use case of high demand storage (mainly metrics and job results).
- Extension of the auto-exclusion feature that has been tested in production with ATLAS and proved effective, to the CMS and LHCb experiments.
- Decommission of Job Robot for CMS, switching to HammerCloud as the main Grid testing service for CMS along with ATLAS.
- Optimization and generalization of submission mechanisms to reduce load.
- Deeper integration with LHCb DIRAC and further deployment for LHCb Grid testing.

#### 3.1.2 CRAB and CRAB Analysis Server

The CMS Remote Analysis Builder (CRAB) [CRAB] was the first analysis tool in CMS to aid users in configuring CMS applications for distributed use [CMSDA], by discovering the location of remote datasets and submitting jobs to the Grid infrastructure. CRAB has progressed from a limited initial prototype nearly 5 years ago to a fully validated system that was critical to the production of published physics results. During the 2010 started the reengineer process aiming both at the consolidation of the core code and at improving the maintenance as well as the sustainability of the tool. Namely the reengineered tool is CRAB3 and currently a beta version is being available for the testing phase.

The asynchronous stage-out tool for user produced data handling is available and a specific monitoring has been developed to make easier the related operations. A thin client has been developed and is able to communicate with the server through the web-services based on a REST interface. The client prototype is currently supporting the 4 basic functionalities: creation, submission, check status, output. A first version of the User Sandbox handling logic has been implemented too.

The main items of development for the next six months are:

- Consolidation of the client and the RESTful interface, implementing additional functionalities: kill, postMortem, report, manual job resubmission etc.
- Final integration with the AsyncStageOut tool particularly refining the CRABServer-AsyncStageOut communication channels.
- Definition and implementation of the logic for the automatic job resubmission.
- Implementation of the user produced data publication into the CMS Data Management System. (Data Bookkeeping System).
- Implementation of the support for the good/bad data filtering by defining Luminosity Masks which are intended to be applied at the data sets.

Longer term goals items are:

- Enabling the support for running bare-root code decoupled by the CMS specific framework (CMSSW).
- Implementation of the support for MonteCarlo end-user private production.
- Development of BossAir plugins for local scheduler support i.e. LSF, condor, PBS, sge.
- Feedback driven tuning, especially at the client level since this is what the user will interact with.

### 3.1.3 Data Management

#### 3.1.3.1 ATLAS Distributed Data Management

The ATLAS Distributed Data Management (DDM) is the project built on top of the WLCG middleware and is responsible for the replication, access and bookkeeping of the multi-Petabyte ATLAS data across the Grid while enforcing the policies defined in the ATLAS Computing Model. The ATLAS DDM project is currently in a mature state and most of the present work is targeted towards the general consolidation of the services, maintenance and support to operations. Existing requests include:

- Improving the source selection for data (e.g. rules to stage data from tape and selection of nearest source)
- Optimized handling of intermediate copies when transferring data between not connected sites

Another area of work that started in the previous year and still remains to be completed is the adaptation of DDM Site Services (the set of agents responsible for data discovery and placement) for Tier3 analysis sites. DDM relies on FTS, LFC and SRM which are all supported in EMI. While these sites are allowed to have local control of their own resources, it is of general interest to have a unified solution to make the data available to them. As a consequence, DDM Site Services are being extended to respect particular data management requirements of these sites and following items have to be finished:

- Testing of new plug-in libraries for the file lookup and transfer job submission using GridFTP (without SRM)
- Integrate and validate complete data placement workflow to Tier3s
- Collect Tier3 access data in DDM information system
- Setup independent Site Services instance serving Tier3s
- Optimize the usage of the new libraries (e.g. parallelization and bulking) after first operational experiences.



### 3.1.3.2 GridGridData popularity and automatic site cleaning

Following first experience with real LHC data taking, production and analysis it has become clear that some of the assumptions behind the current model are not strictly valid. For example, the “pre-placement” of data from Tier1 sites to Tier2s for analysis has shown that a large fraction of the data is never accessed. CMS is working on more dynamic data placement strategies that can result in better resource utilization (potentially lower network traffic as only the needed data is transferred and improved storage management through smaller numbers of data copies).

With this purpose, CMS has recently developed its popularity framework that holds the information about which data is used for analysis at the different Grid sites. The future data distribution model can build upon this framework and replicate further the popular data to facilitate its analysis by the physics community.

On the other hand, unpopular replicas can be reduced when a site is running low on storage space or physics-groups are using more space than originally allocated. For the latter, the automatic site cleaning system, in production for ATLAS since mid 2010 and based on popularity measures is now being adapted for the CMS experiment. The plan for the next months is to fully commission this cleaning system in order to use the system in production.

- In a first stage, the cleaning agent will run on a daily basis and suggest data that can be deleted to the operations team and physics groups, which will have to validate the decisions.
- Implement protections and policies needed to fully address the specific needs of the experiment.
- Evaluate the possibility of implementing a monitoring framework to display the actions and information handled by the cleaning system.
- As a last step, enable the automatic deletion.

### 3.1.4 Persistency Frame-work

The Persistency Framework [PERS-FRAME] is one of the projects set up within the LCG Application Area to provide common software solutions to the LHC experiments at CERN. It consists of three software packages (POOL, CORAL and COOL) that are used by three of the LHC experiments (ATLAS, CMS, LHCb) for storing and accessing several different types of scientific data. CORAL is a generic abstraction layer with an SQL-free API for accessing data stored using relational database technologies. POOL is a generic hybrid technology store for C++ objects and object collections, using a mixture of streaming and relational technologies. COOL provides specific components to handle the time variation and versioning of the conditions data of the LHC experiments.

Plans for the second half of 2011 and 2012 include:

- User support, service operation and software maintenance. These tasks require a large support and development effort as they include, for instance, the preparation of regular software releases (one per month on average, following explicit requests of the experiments for urgent bug fixes, functionality enhancements or upgrades in external software dependencies such as Boost or ROOT) and the follow-up of service operation incidents (which normally result in bug fixes in the Persistency Framework code, but often end up in the need for a more global troubleshooting, involving for instance the Oracle server software or the Grid middleware).
- Development and improvement of new and recently added functionalities such as the capability of the CORAL software to restore the database connection and session after a network glitch, or the review and extension of CORAL monitoring for the CoralServer and Frontier components.
- R&D to evaluate new technologies relevant to data access optimization.
- Review of the support model for POOL, CORAL and COOL with the relevant stakeholders. For POOL this is likely to result in the transfer of support responsibility to the ATLAS experiment, as LHCb recently announced that it will eventually stop using this software component.

### 3.1.5 DIRAC

DIRAC system provides a complete solution for using the distributed computing resources of the LHCb experiment. DIRAC has a complete framework for data production and analysis, including workload management, data management, monitoring and accounting (more details have been given in document [MS610]). One of its most important components is the Data Management System (DMS), whose support in EGI-InSPIRE project started in October 2010. During the past months, significant progress has been done in the improvement and development of the system, mainly in the area of data consistency between storage elements (SEs) and file catalogues and the accounting of space usage at Grid sites.

The main lines of development for the near future are the following:

- Continue the activity of consistency checks between file catalogues and SEs, following a general strategy that will benefit not only LHCb, but also other VOs. The objective of this work is twofold: while on one side it ensures that all the content of the SEs is actually registered in the catalogues, thus avoiding a waste of disk space, on the other hand it also spots any data in the file catalogue that is missing from the physical storage. The fact that Grid SEs and the file catalogue are completely decoupled makes this type of checks necessary for keeping the system consistent.
- LHCb Computing Model foresees to keep only the two latest versions of (re)processed real data on disk. However, it is important to keep datasets that were used for a very long time in order to produce published physics results. This requires a tape archiving mechanism, whose implementation is planned for next year.
- Integration of new systems to provide a sustainable usage of Grid storage resources, in particular to monitor the usage of datasets over the Grid storage systems, provide an efficient policy to remove the least used data, create further replicas of the most used datasets and optimize the use of storage resources.
- Improvement of the system to keep accounting of the historical storage resources usage. Such system has already been developed and put in production, though some optimization is needed to improve its performance and reduce the execution time.
- Adapt when necessary the DIRAC DMS to changes in the data management middleware (FTS, SRM, LFC etc...)
- Provide general support for LHCb data management operations when needed.

### 3.2 *GridGridServices for the Life Science Virtual Research Community*

To coordinate their efforts and sustain their activity, members from the Life Science community self-organized into the project-independent “Life Sciences Grid Community” (LSGC) over the first period of the EGI-InSPIRE project. As described in MS611 [MS611], the Life Sciences HUC contributes to the LSGC effort to organize the community and deliver services. In this context, several services to the HUC users have been planned. Details can be found on the LSGC wiki [LSGC-WIKI]:

- Mailing lists addressing each affiliated VO and the whole VRC have been set up and are updated daily to ensure communication within the community.
- A redundant VOMS server has been provisioned for experimenting fail-safe procedures in case of main VOMS server downtime. Its use in production will shield end-users from unavailability of this critical service.
- A HUC support service is delivered. A technical team of expert users has been set up. It addresses the difficulties reported by users on the VRC mailing lists or through GGUS. It also pro-actively monitors the infrastructure health at a VO level, to identify the problems occurring. The critical services (VOMS, LFC) and the SEs (that are cause of frequent job failures for various reasons) are periodically checked. The CEs will soon be similarly monitored. The French NGI deploys a

dedicated Nagios server<sup>4</sup> for that purpose. Bi-monthly phone conferences are scheduled and shifts are organized to ensure that there is always a team on duty tackling the problems arising.

In the coming year, it is planned to improve the support of the HUC through several actions:

- September 2011 (PM17): Use of the redundant VOMS server in production, as soon as the fail-safe procedures are completely set-up.
- December 2011 (PM20): Similarly, provision of a redundant LFC server, as this service is also a critical single point of failure.

June 2012 (PM26): Development of a HUC user's database and management tools to assist VO administrators in their daily task. The designs of the database schema and user registration state diagrams are complete. This service will interface to Virtual Organization Membership Service (VOMS) servers as well as the EGI applications database, to avoid replicating existing information. It will complement the VOMS and applications database with extra-information on the users and their affiliations.

### ***3.3 Services for the Astronomy & Astrophysics Virtual Research Community***

The activity related to services for A&A community focused on the following topics: visualization tools and in particular VisIVO; parallel processing on the Grid: MPI and CUDA; Grid-HPC interactivity; access to databases and integration with the Virtual Observatory.

An internal work-plan was prepared to identify in detail the sub-tasks and the involved resources.

During the first year, the most relevant activity concerned the porting of VisIVO (Visualization Interface to the Virtual Observatory) in Grid. The first step related to this activity consisted in enabling the usage of VisIVO directly within a code during the production phase. A software layer therefore has been developed for this purpose using the internal arrays, without the need of producing intermediate files. A library of VisIVO was designed and implemented for this. The first issue of the Gridified version of VisIVO was released in April 2011 (MS608 of EGI-InSPIRE). Recently (starting from Q5) the activity focused on the study and on the porting of the MPI-enabled version of VisIVO in gLite and on the integration of VisIVO on Grid nodes where GPUs are available.

For what concerns Grid and HPC, the collection of significant A&A use-cases and test-beds is in progress in the context of a coordinated EGI.eu and NGIs effort within the A&A VRC; the evaluation of tools and services currently in place to integrate/interface Grid infrastructures and databases as well is in progress. In this context the following A&A applications have been identified: FLY (INAF-OACT Cosmological code) and Gadget + Flash, the most common cosmological codes in Astrophysics.

The integration in Grid of the BaSTI (A Bag of Stellar Tracks and Isochrones) Astronomical Database and its feeding FRANEC code took place as well. A web portal has been developed to make easier the submission of the FARANEC code in Grid.

A considerable amount of time was also dedicated to the coordination of the European A&A community in EGI in order to stimulate and to foster the requirements gathering process to be fed to EGI.

According to plans for the second year of EGI-InSPIRE, the activities related to VisIVO, to Grid and Supercomputing interactions, and the work on BaSTI and FRANEC will continue in the next months. A consistent part of the activity will be dedicated to the harvesting of requirements, use-cases and test-beds to be provided to EGI. To boost this activity and overcome the current standstill situation a strong coordination of the A&A community is mandatory in the next months. For this coordination activity the forthcoming events (especially the EGI Technical Forum in September 2011) are of

---

<sup>4</sup> Biomed Nagios server, <https://Grid04.lal.in2p3.fr/nagios>

utmost importance. In addition, a specific workshop for the A&A HUC, is under preparation. This event will take place in Paris in November 2011, in the framework of the Astronomical Data Analysis Software and Systems (ADASS) XXI Conference. The main goals of the workshop consists in strengthening the A&A HUC activities in EGI and in fostering the formal endorsement of the A&A VRC by EGI; to this end, the main objective of the workshop is to make the A&A community in the right position to sign the proposed MoU (or at least the LoI) with EGI.

### ***3.4 Services for the Earth Science Virtual Research Community***

The Earth Science community is working on a few smaller development projects regarding the access to Earth Science data infrastructures, mainly the exploitation of the GENESI-DR infrastructure and the Earth System Grid (ESG). For EGI users, the GENESI-DR infrastructure provides a standardized data discovery interface based on OpenSearch and metadata standards for a federation of data repositories. While in the European project behind it (GENESI-DEC) focuses on a Web portal as an interactive entrance point, the usage on EGI requires a non-interactive, bulk oriented, tool. The work on such a command line client that facilitates the usage in Grid jobs is ongoing. The work in the next months will involve the inclusion of a data transfer component that downloads the bulk of resources found through the discovery process. This is required as the discovered data sets are not accessible through a unified channel, but depending on the data set instead through different access protocols such as HTTP, GridFTP or others.

Regarding interoperability with the Earth System Grid, a testing scenario is developed. It is based on an application from IPSL which uses climate model data stored in the ESG belonging to the fifth phase of the Climate Model Intercomparison Project (CMIP5). The scenario involves software development on three different components: a CMIP5 data transfer tool, a credential testbed and the application itself. The data transfer tool is a standalone Python command line application that facilitates the access to ESG data from the command line. Based on input parameters such as model and experiment, the tool looks up the corresponding data node and downloads the data. For the next months it is planned to add additional data selection options, bandwidth usage statistics and more configuration options. The credential testbed is used for prototyping solutions to the challenge of different authentication schemes on EGI and ESG. Currently separate credentials are required for EGI and ESG, which imposes problems on non-interactive usage in Grid jobs. The testbed will be extended according to the results of the ongoing discussion with the ESGF. The application will be modified accordingly to utilize the credential testbed and data transfer program running on the EGI.

The results of the development efforts will be made available on [euearthscienceGrid.org](http://euearthscienceGrid.org).

## 4 CONCLUSIONS

All the Heavy Users Communities included in the SA3 work-package have contributed to this report on the HUC Software Roadmap with information on the status and planning of their services and tools. New versions of this document are produced every 6 months, starting from [MS602] in Project Month 4 and [MS606] in Project Month 10. The communication between the different communities has started and is growing – this is a visible and concrete deliverable of the project where it is clear that the funding model and goals of the project are succeeding in motivating common tools and services, even if at the architectural but sometimes also at the implementation and deployment level; the planning for their software of potentially more general interest, has been exposed to the other communities of heavy users and to the general users communities. As described in detail in the Deliverable D6.2 on Sustainability plans for the HUCs [D6.2], a variety of common areas between the HUCs have already been identified, with solutions being adopted by multiple VOs and considered also by others. Concrete examples include the HammerCloud Stress Testing System for Distributed Analysis [HC], now adopted by ATLAS, CMS and LHCb as well as the more recent work on dynamic data placement and caching (aka Data Popularity). Users wishing to know more of the capabilities of these services should consult D6.4 “Capabilities Offered by the HUCs to Other Communities” [D6.4], as well as MS609 “HUC Contact Points and Support Model” [MS609]. The achievements of the work package in the first year of the project are detailed in D6.3 “Annual Report on the HUC Tools and Services” [D6.3].

As defined by the mandate of this work package, the feedback from the different kind of users, the sites and the software developers will continue to be an important next step towards the long term planning of these services and tools for sustainability.

## 5 REFERENCES

|               |                                                                                                                                                                                                              |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>MS105</b>  | MS105 Quarterly Report 1:May-July 2010:<br><a href="https://documents.egi.eu/public/ShowDocument?docid=156">https://documents.egi.eu/public/ShowDocument?docid=156</a>                                       |
| <b>MS106</b>  | MS106 Quarterly Report 2:August-October 2010:<br><a href="https://documents.egi.eu/public/ShowDocument?docid=248">https://documents.egi.eu/public/ShowDocument?docid=248</a>                                 |
| <b>MS107</b>  | MS107 Quarterly Report 3:November 2010-January 2011:<br><a href="https://documents.egi.eu/public/ShowDocument?docid=361">https://documents.egi.eu/public/ShowDocument?docid=361</a>                          |
| <b>MS601</b>  | MS601 – HUC Contact points and the support model:<br><a href="https://documents.egi.eu/document/91">https://documents.egi.eu/document/91</a>                                                                 |
| <b>MS602</b>  | MS602 – HUC Software Roadmap: <a href="https://documents.egi.eu/document/230">https://documents.egi.eu/document/230</a>                                                                                      |
| <b>MS603</b>  | MS603 – Services for High Energy Physics:<br><a href="https://documents.egi.eu/document/160">https://documents.egi.eu/document/160</a>                                                                       |
| <b>MS604</b>  | MS604 – Services for the Life Science Community:<br><a href="https://documents.egi.eu/document/236">https://documents.egi.eu/document/236</a>                                                                |
| <b>MS605</b>  | MS605 – Training and dissemination event for all shared services and the other tasks within the activity: <a href="https://documents.egi.eu/document/326">https://documents.egi.eu/document/326</a>          |
| <b>MS606</b>  | MS606 – HUC Software Roadmap:<br><a href="https://documents.egi.eu/public/ShowDocument?docid=310">https://documents.egi.eu/public/ShowDocument?docid=310</a>                                                 |
| <b>MS607</b>  | MS607 – Hydra service deployment: <a href="https://documents.egi.eu/document/327">https://documents.egi.eu/document/327</a>                                                                                  |
| <b>MS608</b>  | MS608 – Integration of the VisIVO server with the production infrastructure:<br><a href="https://documents.egi.eu/document/328">https://documents.egi.eu/document/328</a>                                    |
| <b>MS609</b>  | MS609 – <a href="https://documents.egi.eu/public/ShowDocument?docid=419">https://documents.egi.eu/public/ShowDocument?docid=419</a>                                                                          |
| <b>MS610</b>  | MS610 – Services for High Energy Physics:<br><a href="https://documents.egi.eu/secure/ShowDocument?docid=540">https://documents.egi.eu/secure/ShowDocument?docid=540</a>                                     |
| <b>MS611</b>  | MS611 – Services for the Life Science Community:<br><a href="https://documents.egi.eu/secure/ShowDocument?docid=683">https://documents.egi.eu/secure/ShowDocument?docid=683</a>                              |
| <b>D6.1</b>   | D6.1 – Capabilities Offered by the Heavy User Communities:<br><a href="https://documents.egi.eu/public/ShowDocument?docid=154">https://documents.egi.eu/public/ShowDocument?docid=154</a>                    |
| <b>D6.2</b>   | D6.2 – Sustainability plans for the activities of the Heavy User Communities:<br><a href="https://documents.egi.eu/public/ShowDocument?docid=309">https://documents.egi.eu/public/ShowDocument?docid=309</a> |
| <b>D6.3</b>   | D6.3 – Annual Report on the Tools and Services of the Heavy User Communities: <a href="https://documents.egi.eu/document/312">https://documents.egi.eu/document/312</a>                                      |
| <b>D6.4</b>   | D6.4 - <a href="https://documents.egi.eu/secure/ShowDocument?docid=472">https://documents.egi.eu/secure/ShowDocument?docid=472</a>                                                                           |
| <b>MONARC</b> | MONARC Technical Notes:<br><a href="http://www.cern.ch/MsONARC/docs/monarc_docs.html">http://www.cern.ch/MsONARC/docs/monarc_docs.html</a>                                                                   |



|                    |                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    |                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>DASHBOARD</b>   | Experiment Dashboard for Monitoring of the LHC Distributed Computing Systems;<br>J. Andreeva et al, Computing in High Energy and Nuclear Physics (CHEP'10), Taipei, Taiwan (2010)                                                                                                                                                                                                                                          |
| <b>GANGA</b>       | Ganga: a tool for computational-task management and easy access to Grid resources; <i>Computer Physics Communications, Volume 180, Issue 11, (2009)</i>                                                                                                                                                                                                                                                                    |
| <b>HAMMERCLOUD</b> | HammerCloud: A Stress Testing System for Distributed Analysis; Daniel C. van der Ster et al, Computing in High Energy and Nuclear Physics (CHEP'10), Taipei, Taiwan (2010)                                                                                                                                                                                                                                                 |
| <b>GREIC</b>       | <a href="http://www.grelc.unile.it/">http://www.grelc.unile.it/</a>                                                                                                                                                                                                                                                                                                                                                        |
| <b>HYDRA</b>       | <a href="https://twiki.cern.ch/twiki/bin/view/EGEE/DMEDS">https://twiki.cern.ch/twiki/bin/view/EGEE/DMEDS</a>                                                                                                                                                                                                                                                                                                              |
| <b>KEPLER</b>      | <a href="https://kepler-project.org/">https://kepler-project.org/</a>                                                                                                                                                                                                                                                                                                                                                      |
| <b>GRIDWAY</b>     | <a href="http://www.Gridway.org/doku.php?id=start">http://www.Gridway.org/doku.php?id=start</a>                                                                                                                                                                                                                                                                                                                            |
| <b>SOMA2</b>       | <a href="http://www.csc.fi/soma">http://www.csc.fi/soma</a><br><br>Kinnunen, T., Nyrönen, T., Lehtovuori, P., SOMA2 - Open Source Framework for Molecular Modelling Workflows, <i>Chemistry Central Journal</i> , 2(Suppl 1):P4 (2008).<br><br>Lehtovuori, P. Nyrönen, T., SOMA - Workflow for Small Molecule Property Calculations on a Multiplatform Computing Grid, <i>J. Chem. Inf. Model.</i> , 46(2) (2006) 620-625. |
| <b>MPI</b>         | <a href="http://egee-uig.web.cern.ch/egee-uig/beta_pages/MPIJobs.html">http://egee-uig.web.cern.ch/egee-uig/beta_pages/MPIJobs.html</a>                                                                                                                                                                                                                                                                                    |
| <b>WIKI</b>        | <a href="https://wiki.egi.eu/wiki/SA3:_Services_for_the_Heavy_User_Community">https://wiki.egi.eu/wiki/SA3:_Services_for_the_Heavy_User_Community</a>                                                                                                                                                                                                                                                                      |
| <b>CRAB</b>        | CRAB: A CMS Application for Distributed Analysis<br>Nuclear Science, IEEE Transactions on Volume 56, Issue, Part 2, Oct. 2009 pp:2850 - 2858                                                                                                                                                                                                                                                                               |
| <b>CMSDA</b>       | Distributed Analysis in CMS;<br><i>Journal of Grid Computing</i> , Vol.8, Number 2, 159-179, (2010) doi: 10.1007/s10723-010-9152-1                                                                                                                                                                                                                                                                                         |
| <b>CMSAOPS</b>     | CMS distributed analysis infrastructure and operations: experience with the first LHC data;<br>E. Vaandering et al, Computing in High Energy and Nuclear Physics (CHEP'10), Taipei, Taiwan (2010)                                                                                                                                                                                                                          |
| <b>PHEDEX</b>      | Software agents in data and workflow management;<br>T. Barrass et al, Computing in High Energy and Nuclear Physics (CHEP'04), Interlaken, Switzerland, 2004                                                                                                                                                                                                                                                                |
| <b>PERS-FRAME</b>  | <a href="https://twiki.cern.ch/twiki/bin/view/Persistency">https://twiki.cern.ch/twiki/bin/view/Persistency</a>                                                                                                                                                                                                                                                                                                            |





|           |                                                                                                        |
|-----------|--------------------------------------------------------------------------------------------------------|
| LSGC-WIKI | LSGC wiki, <a href="http://wiki.healthGrid.org/LSVRC:Index">http://wiki.healthGrid.org/LSVRC:Index</a> |
|-----------|--------------------------------------------------------------------------------------------------------|