Life Science strategic cluster workplan

Goals

The Life Sciences (LS) cluster comprises three main research areas currently using EGEE: medical imaging, bioinformatics and drug discovery. The targeted communities include tens of thousands of researchers in Europe. Presently, with more than 30 applications deployed and being ported, the domains had more than 200,000 jobs executed per month in 2007.

The main objective of the cluster is to increase the impact of EGEE-III in these strategic communities which have similar problems regarding the integration of existing data on the Grid, secured access to that data, orchestration of complex tasks on the Grid (workflow), the use of licensed software and the availability of high-level interfaces like portals and Web services that hide the complexity and the technical tricks of grid infrastructure-. The primary theme for this cluster will be advanced data management focusing primarily on secured access, encryption and data flow controls.

The work within the biomedical strategic cluster covers the three areas: specific support, development/evaluation of software, evolution of gLite and intensive testing. The subtasks identified in the LS strategic cluster are the following:

- support for selected services of particular interest to the biomed community : AMGA, MOTEUR and the Medical Data Manger (686PM)
- preparation of the migration to EGI in the life sciences sector (2018PM)
- application porting in the fields of medical imaging, bioinformatics and drug discovery (18580PM)

Partners

Partner name	Country	Person-Months
ASGC	Taïwan	24
CNR-ITB	Italy	18
CNRS	France	90
CNU	Korea	84
KISTI	Korea	390
UPV	Spain	18
TOTAL		27364 PM

Description of subtasks

Subtask 1: support for selected services (686PM)

The cluster will provide support to the discipline for using two advanced data management services already available: AMGA for metadata management, MOTEUR for data-oriented workflow enactment and the Medical Data Manager for medical image servers. The cluster will complement the present support provided almost exclusively by the developer teams that are now facing a growing pressure with the increasing adoption of their services. Cluster partners will share their expertise on AMGA and MOTEUR with the members of the NA4 support team to further increase user support efficiency.

Partner name	Country	Contribution to sub task	Person-Months
ASGC	Taiwan		0
CNR-ITB	Italy	Support to workflow	2̶0
CNRS (I3S & CREATIS)	France	Support to MOTEUR	33
CNU	Korea		0
KISTI	Korea	Support to AMGA	21
UPV	Spain	Support to AMGA	12
TOTAL			68̶6̶PM

Within the framework of this subtask, the cluster foresees interactions with the teams developing the Taverna workflow engine and the VLe VBrowser. The cluster cannot commit to handle these developments on its own but ̶it will offer assistance and guidance regarding the porting of these services to gLite. The completion of the development related to these external tools will depend on the involvement of the teams developing these tools.

Subtask 2: preparation of the migration to EGI in the life sciences sector (20̶1̶8̶PM)

Migration to the European Grid Initiative requires preparation. Indeed, the biomedical community has been significantly involved in the grid infrastructure projects funded by the European Commission since the early days of DataGrid. If the particle physics community has already its own international production grid, the LHC Computing Grid, the different disciplines represented in this cluster are not yet at this stage. Some of them like the life sciences community are presently running design studies for the deployment of pan-european infrastructures which could take advantage of the grid technology.
A first goal of this subtask is to contribute to these design studies wherever possible to promote the use of grids and particularly of EGI. A second goal is to contribute information, use cases and requirements to the EGI design study in order to prepare the transition after the end of EGEE-III.

Partner name	Country	Contribution to sub task	Person-Months
ASGC	Taiwan		0
CNR-ITB	Italy	Interaction with BBMRI, EGI, ELIXIR and other design studies	2
CNRS (LPC Clermont-Fd & IBCP)	France	Interaction with EGI, ELIXIR and other European initiatives (VPH)	18
CNU	Korea		0
KISTI	Korea		0
UPV	Spain		0
TOTAL			20PM

In relation to this subtask, members of the Life Sciences cluster will also pursue their involvement in a number of bodies and groups to express the requirements and defend the interest of the biomedical community.

Subtask 3: support to application porting in the field of bioinformatics, medical imaging and drug discovery (1850PM)

Application porting is still an important activity of the Life Sciences cluster. The largest contribution to this subtask comes from the Asian partners. With more than 30 applications currently running on its production infrastructure, EGEE is most certainly the largest biomedical infrastructure in the world. However, this vitality is still fragile and it is particularly important to keep steering the deployment of new applications, especially in relation to the availability of new services which may require debugging or the feedback of advanced users. A few applications which have reached a certain level of maturity and visibility like the MR image simulator, Therapeutic Irradiation Simulator or WISDOM will be further supported during the lifetime of the project. They will be used to test new services and also to study the migration to EGI.

Partner name	Country	Contribution to sub task	Person-Months
ASGC	Taiwan	Drug discovery	24
CNR-ITB	Italy	Bioinformatics	148
CNRS (CREATIS, I3S, LPC Clermont-Fd & IBCP)	France	Bioinformatics, Drug Discovery, Medical Imaging	39
CNU	Korea	Drug discovery	84
KISTI	Korea	Bioinformatics	189
UPV	Spain	Medical Imaging	6
TOTAL			1850PM

Metrics

We propose the following metrics to evaluate the activity of the Life Sciences cluster during the project lifetime:
- Number of applications ported on the EGEE Grid infrastructure, and also the improvement in term of efficiency of the application ported on the grid
- Number of European infrastructures design studies contacted, and involvement and level of participation in these design studies
- Number of AMGA and MOTEUR users supported directly or through the user support team
- Number of tools, services and applications successfully tested.
- Number of applications in the life sciences sector
- Number of users in the Virtual Organizations of the life sciences sector
- Resources consumed by the Virtual Organizations in the life sciences sector

Procedures and tools

The EGEE-III life sciences cluster builds upon the collaboration between the partners which has developed in the previous years. The partners share a commitment to open source. Regarding the procedure to accept an application on the biomed VO, there is a clear preference to keep the present approach which is very flexible and informal.

In the medical imaging sector, attempts will be made to deploy a medical image analysis tools repository. To reach the user community, this repository needs to index image processing tools using a high level representation schema and relevant information. This will should contribute in increasing the grid infrastructure usability, by providing to new users not only an

infrastructure but also immediately accessible software components that will help the design of new applications.

Communication plan

Foreseen mechanisms for internal communication are the ones used so far: mailing lists, phone conferences, wiki.
Contribution to other EGEE bodies like TCG is also planned. For example, the Life Science cluster activity has taken a leading role, and will continue in this way, in the PORTAL workgroup (TCG) during EGEE-II about related topics to "Portal access to the Grid", which is a strong requirement from several Life Science applications.

External communication includes Web/wiki, iSTWG, conferences (from grid-oriented to end users-oriented: HealthGrid, MICCAI (a grid specific workshop organized this year), RSNA, bioinformatics workshop at CCGrid.

Tools

Regarding the question of setting up a specific development server for the cluster, it is not clear if the developers of specific services will agree to externalize their development base on a centralized biomed server. A constellation of development servers/tools is rather expected.