Subject:	**ANSWER TO EAC FEEDBACK, FOLLOWING EGEE ATHENS CONFERENCE**

Author(s):	**Marc-Elian Bégin on behalf of the PO and PEB**

Distribution	**PO, PEB**

This note is the answer to the EAC feedback provided to the project at the end of the EGEE Athens Conference. The comments have been numbered, for ease of referencing, and the answers or replies have been highlighted in blue.

1. APPLICATIONS

1. May want to refocus how "other applications" are to be identified, managed and engaged. Seems to be a tendency to adopt the "build it and they will come" approach, when it is really a very intensive effort as it requires change in practice and approach.

Reply: To acknowledge the fact that EGEE cannot provide full support to the growing number of Grid applications, for EGEE-II we are proposing to rely more heavily on related application Grid projects, while continuing our important role as incubator for new communities and applications. This process will allow a large number of new applications to be tested on the various EGEE Grid infrastructures, and only when the applications have demonstrated the added value of the Grid will they be promoted to "supported" applications. This new process addresses both the need for a lightweight process of evaluation (both for EGEE and the new applications), while keeping the support structure affordable to ensure the required quality of services and support expected from EGEE.

In EGEE-II, we will continue to support the *gridification* and deployment of diverse scientific communities, including: *High Energy Physics*, *Biomedicine*, *Earth sciences*, *Astrophysics*, *Computational Chemistry* and *Geophysics* (which include the industrial application, EGEODE), as a continuation of EGEE. Further, a number of "related project" and new Grid communities, such as *Fusion*, have also stated their intention to deploy applications on the EGEE infrastructure. See selection process in the reply to comment 5.

2. The role and function of digital library activities; still seems to be mostly like smoke and mirrors - they need a list of specific deliverables and metrics in 3-6 months or it will detract from success of EGEE - this is not being critical of digital libraries, but is a comment that they need to become much more focused on 1 or 2 small projects or applications that they can demo. They also need to make a stronger case on how they can leverage EGEE or why they even need a grid.

Reply: Two exemplary and complementary application scenarios, which will assist the identification of user requirements and validate the DILIGENT framework, have been chosen for inclusion into the DILIGENT project: the ImpECt and the ARTE scenarios.

The goal of the ImpECt scenario is to improve accessibility, interoperability and usability of environmental data, models, tools, algorithms and instruments integrating the distributed data sources with specialized data handling services.

ImpECt (Implementation of Environmental Conventions) includes leading actors in the environmental sector, and is represented by the European Space Agency.

The goal of the ARTE scenario is to stimulate collaborative, multidisciplinary scientific research, to ease multimedia artifact construction and to improve support for education.

The ARTE community is represented in DILIGENT by the Scuola Normale Superiore (SNS), one of several observers contributing rich archives of texts and images. Audio-visual content is being provided by the Italian National Broadcasting RAI Educational.

[Meb] Still not capturing the Grid usecase (inputs expected from Florida)

3. CrossGrid Applications
    o   These should be ported to gLite

Reply: The EU CrossGrid project (IST-2001-32243) has approached EGEE in order to set up a clear relation between the projects and use synergies for the sustainable uptake of CrossGrid technology and the CrossGrid testbed. At the third EGEE conference, colleagues from CrossGrid received the prize for the best demo, based on the selection of the EAC. Following this, EGEE has invited respective CrossGrid partners to cooperate on a series of exemplary components: flooding crisis application, *GridBench* performance benchmark, and the *glogin* tool.

Together with the Slovak Academy of Sciences (II-SAS), a plan for porting the flooding crisis application from CrossGrid to gLite has been set up. The implementation is currently ongoing and is expected to conclude within the next few months. Similarly, EGEE and CrossGrid have studied possibilities to port the CrossGrid benchmarking tool GridBench to gLite. In this case, a detailed work plan is currently being established, which will be carried out in the remainder of EGEE and EGEE-II.

The glogin tool, enables tunnelling of interactive communication between nodes on the Grid and the user's desktop, which was originally developed in the EU CrossGrid project and is now applied in a series of other grids (e.g. Austrian Grid), has been ported to gLite. More detailed investigations about its utilization for applications (e.g. gPMT3D) as well as a tool for grid site administration are ongoing.

It is a general problem to fund this type of effort. In the case of GridBench and glogin, the porting was performed through additional non-EU budget.

It is likely, and this will probably be a measure of success in many ways, that more services, tools, plug-ins and component will be made available from within and outside the project, building on the gLite foundation middleware. We need to understand how deal with these "middleware extensions" to gLite, without necessarily becoming a part of the gLite middleware itself. We could think of cataloguing these extensions and have them listed somewhere on our website and repository. We also need to think about a certification process, or something similar. Finally, we need to document this process in the TA, in time for the negotiation with the EU. An option could be a repository - such as gLite/sourceforge and/or built on ETICS - or even offloading it to an "OMII-Europe"-like project (??!).

4. Applications need more focus on Grid potential

Reply: We have increased our focus on identifying the key and unique advantages the Grid offers applications in several ways:

-   Production of a project video, including several interview extracts from users, talking about the benefits the Grid is providing to their research programme

-   Several specific application focused press-releases (see here for details)

Further, we are also expanding the user level services of the Grid with specific activities, such as:

-   the service challenge deployed by LHC experiments addressing high throughput distributed data storage

-   workflow tools like TAVERNA and DAGMAN being interfaced with EGEE middleware opening the perspective of deploying more complex tasks

- data management services targeted at specific domains, such as the DICOM/SRM interface, giving access from the Grid to medical raw data

5. Define an application selection process
   o Good Grid and EGEE project ambassadors

Reply: As mentioned in our reply to comment 1 above, the selection process has been updated for EGEE-II, and we are working on a trial period within the current phase in order to evaluate its efficiency. This new process will accommodate both sources of Grid applications and user community maturity by allowing small and informal evaluation of the EGEE Grid infrastructures, via a dedicated local resource centre, while allowing more organised projects and communities to enter into negotiation directly with the project.

6. EGEE should be cautious to support new applications if the support is costly compared to the gain in overall experience and attractiveness for important new classes of grid usage.

Reply: We acknowledge this risk. As mentioned in the reply to comment 5 above, for EGEE-II we are proposing new processes for application incubation and support. Further, no EGAAP call for new application domains has been issued since Athens.

7. Forward Look
   o Need an EGEE "Grand Challenge"

Reply: During the summer of 2005, the biomedical pilot application community ran a Data Challenge on Drug Discovery. This challenge used 1000 CPUs over 15 countries, cumulating in over 80 years of equivalent CPU usage. For more details see the press release.

8. It may be useful to emphasize VO "hero" goals which clearly identify specific accomplishments and/or identities for various VOs. This helps with buy-in at the local/regional level and gives them some unique identity with the larger EGEE effort.

Reply: See the reply to comment 7. More press releases have been issued on specific application achievements (see the press release).

2. GLITE

9. Full Project Sign-up
   o LHC sign-up to using gLite

Reply: The LHC experiments are putting in place task forces to address specific issues related to the computing infrastructure, which will provide EGEE with a simpler interlocutor. Meanwhile, the Pre-Production Service is being used by the LHC experiments to evaluate the EGEE middleware and provide further feedback. Finally, the CERN Director General, Robert Aymar, has stated during a recent LCG-POB meeting that he expects LCG and the LHC experiments to use the EGEE middleware. See also reply to comment 11 below.

10. Transition and migration of LCG2 to gLite - clearly significant work still to be done - more homework, integration, coordination, buy-in across the project.

Reply: SA1 has already released in production several gLite services. For example, VOMS, RGMA and FTS (see reply to comment 12 for further details) are already deployed on the production service. Meanwhile, the pre-production service (PPS) now includes gLite 1.3 (and will deploy 1.4 as soon as available). The PPS is now being used by CMS HEP experiment. Further, the DILIGENT project has deployed gLite on a separate testbed, in order to build their own infrastructure. gLite is also deployed on the GILDA testbed and has already been used at a number of training events over the summer.

Further, the architecture document of the EGEE middleware was updated this summer. See reply to comment 13 below for more details.

Here is the latest status of the Grid Foundation Middleware:

- **Security model and infrastructure**

  This comprises all the necessary components to define and enforce policies within VOs as well as those needed by the resource provider (i.e. between VOs). In particular, resource access control, resource access auditing and VO membership management needs to be provided as part of an integrated security infrastructure. In addition to the maintenance and hardening of existing gLite components this work will focus on policy definition and enforcement, dynamic connectivity, auditing, and interoperability with Shibboleth-based authentication and authorisation infrastructures.

- **Computing Element** (CE)

  The Computing Element is a set of services that provide homogeneous, managed, and secure access to heterogeneous, remote computing resources. It provides structured and secure mechanisms to allow higher level services to either directly submit jobs or to dynamically deploy VO specific schedulers. The collaboration with Condor and Globus, started in EGEE, will be continued to converge on common solutions.

- **Storage Element** (SE)

  Similar to the CE, the Storage Element provides homogeneous access to storage resources, including managed data transfer. While relying on externally provided storage systems offering an SRM interface, POSIX-like I/O, and the file transfer service (FTS) need to be provided.

- **Accounting**

  Accounting is concerned with collecting the relevant information locally at the resources and making it available at a global or VO level (in a secure manner) for statistical, billing, or scheduling purposes. The currently available accounting prototypes need to be further hardened and harmonised.

- **Information and monitoring**

  Information published by the various services on the resources they control must be accessible to other services in a dependable and timely manner. The infrastructure must also be able to support the monitoring of resources and allow user-level monitoring. This includes information and monitoring infrastructure, and service discovery. The work started in EGEE needs to be further hardened and interfaces need to be harmonised across different information systems to allow free information flow across different Grid flavours.

11. I did not win confidence that gLite will work well and according to the timetable. There is no decision from LHC operation visible to me that gLite will be the used in a decisive role. There is still the fear on my side that EGEE had to adapt to the LHC requirements only in order to learn at a late point in time, when it has lost other communities, that it is not accepted by CERN.

Reply: The time between releasing software from JRA1 and having it deployed on the production service is measured in months not weeks, and this is especially true when introducing a significant change in the middleware distribution. The LCG baseline services definition (a document capturing the essence of what the LHC experiments require in production from the Grid middleware) matches closely what is being offered by gLite and this has been met with general

acceptance by the LHC community. Further, to reinforce the process of selection of Grid middleware deployed on the production service, we are planning to introduce the Technical Coordination Group (defined in the EGEE-II proposal) as early as November 2005. This group will provide all the major application groups and technical activities a platform to share inputs into scheduling and priorities for middleware re-engineering, integration, testing, certification and deployment.

12. Be more aggressive in deployment

    o Almost too gentle transition phase

Reply: Several gLite services have already been deployed in Production. R-GMA, the information system is used by LCG for monitoring. The new File Transfer Service (FTS) is being used as a cornerstone of the LCG Service Challenges, responsible for the heavy data transfer between CERN and the tier-1 centres. Meanwhile, the full gLite middleware stack is being deployed on the Pre-Production Service, and tested, in preparation for deployment on the Production Service.

13. Still not a cohesive architecture

    o Alien, CREAM, ...

Reply: During the summer of 2005, we have re-issued the architecture document, describing the updated architecture of gLite, integrating lessons learned from testing, using and deploying earlier versions of gLite. This updated architecture is now consistent and coherent, where good ideas from alternative architectures have been integrated in the gLite architecture.

14. Data Management in gLite was criticised strongly.

Reply: Since Athens, the data management solution offered by EGEE with gLite has considerably progressed. gLite now provides a reliable File Transfer and Placement Service. The biomedical community is currently evaluating the Metadata Catalog, an important aspect of data management. The SRM specification has also progressed and gLite is tracking these features, now offering, for example, SRMcpy to allow direct copy between SRM instances.

15. Although not critical for EGEE, data storage and file management, especially at scale and distribution is a problem which must be addressed by EGEE-II. "Data management in different Grid middle wares and applications is in a terrible state" - Marko Niinimaki

Reply: Further to the reply to comment 14, Mass Storage Systems is out-of-scope for EGEE-II.

3. SUSTAINABILITY

16. Not just ready for the long term ...

    o It is time for the long term

Reply: [Extract from EGEE-II Proposal] "It is planned that significant parts of the infrastructure be transferred over to a new organisation already in its planning stage and referred to as the European Grid Organisation (EGO). Uptake of the technology and its maintenance by EGO will ensure an optimum long-term sustainable use of the infrastructure on a European scale, in a way that responds effectively to market forces".

Also, see attached note from the CERN Director General on the longer term vision of the Grid.

17. Now is time for a business model(s) for the sustainability of EGEE infrastructure.

    • Pure commercial

    • Network infrastructure

Reply: [Extract from EGEE-II Proposal] "One of the main goals of EGEE-II is to better integrate the Grid middleware with industrial solutions in both hardware and software. As a means of extending the platforms on which EGEE-II is available, the project has established a close link

with the CERN openlab project". The collaboration with openlab should be an opportunity to explore *Grid Economies.*

Further, the Société Générale de Géophysique has already deployed a first commercial application on the Production Service (more details here), which provides a novel collaboration platform between industry and public research.

We are making sure that the hooks are in place in the "Grid Foundation" layer of the EGEE middleware, such that these services can be built. The *Vision* section of the proposal explains how EGEE-II will bridge the current Grid initiative with a longer term sustainable solution called the "European Grid Organisation". See 18.

18. "Pay for EGEE" Scheme

   o Application that don't want to share

Reply: Before we can offer a full Grid economies scheme, including a "pay as you go" service, we need to build the hooks into the Grid foundation middleware, such that higher level services can reliably be built. While such high-level Grid economies services are out of scope for EGEE-II, we are committed to providing the hooks for other Grid projects, and potentially openlab partners, to explore such important services.

On the other hand, applications that do not or cannot share resources (e.g. for privacy or security reasons) can follow the MammoGrid deployment model: take the EGEE middleware and deploy it on a private Grid. The same approach is planned for Health-e-Child, where in this case they will get training and will be given access to the GILDA testbed in early stages to learn and plan their own deployment. This private deployment can either be a temporary solution, until the Production Service supports better security and encryption, or permanent solution.

4. MISCELLANEOUS

19. Training needs to continue

   o Regional or VO focus

Reply: Training needs to continue to be both regional and VO based. People need training presented in their local languages and at local venues which they can easily access. VOs also need to be supported and will require domain specific training. This has been explicitly addressed in the EGEE-II proposal where the need to focus on each of these approaches is detailed for each of the training tasks. Similarly a specific deliverable relating to VO specific training has been suggested.

   o Material provided centrally

Reply: Material is indeed held centrally in the materials archive at:

http://www.egee.nesc.ac.uk/trgmat/index.html

This is material created by all of the partners and is open for use/re-use by anyone who needs it and re-submission of modified material is encouraged. This archive is also the basis for a more capable eLearning system which will be demonstrated in Pisa, for instance providing collaborative metadata generation facilities.

20. Managing Expectations

   o Marketing

Reply: In order to refocus our high-level public relation message, we have developed a new series of printed and multi-media material. For example, the EGEE brochure and the EGEE short version video have been produced to directly target executives and decision makers.

Further, the project is also careful not to raise expectations of what can be achieved, by maintaining clear, consistent, open and honest communication with all stakeholders (media, users,

new users, staff and so on). The project will continue to ensure EGEE is presented in the most realistic light, whilst at the same time not under selling the real advantages of EGEE to science and beyond.

Meanwhile, in order to better define the expectations from both the EGEE project, new user communities and their application groups, light-weight Memorandums of Understanding (MoU) are being established with each and every supported application. Each MoU is helping to make clear what the users can expect from EGEE and what their engagements are with respect to EGEE.

- o Professional support for the public demos

Reply: In EGEE-II, the company Metaware, which has experience in marketing for decision-makers, business and government, has joined the NA2 activity. Further, User Forums will be held at regular intervals, which should provide the environment to provide more demos to a larger audience and feedback from professionals, such as Metaware.

21. OGSA-Naming involvement to address data naming problem

Reply: EGEE is following the OGSA-NAMING design team but for manpower reasons we do not contribute actively.

22. Service Challenges a good idea

Reply: The Service Challenge has indeed been a good source of focus, which brought LCG and EGEE closer together. The results of these challenges have produced quality services such as the File Transfer Service (FTS), but also provided specific feedback used to harden other services such as R-GMA. EGEE will continue to actively support these types of specific efforts.

23. Quality Assurance Meeting was very quiet. Only the leader spoke. They did some planning and simply talked about the content they were monitoring. Very little on what they would do proactively with this data to improve EGEE.

Reply: Since Athens, several follow-on meetings have been held between technical experts from JRA2, SA1 and JRA1. These meetings lead to a roadmap, providing short-term and longer-term solutions for job monitoring. Further, the EGEE-II proposal specifies several project-level targets, which will also be tracked by a common effort between the different technical activities.

24. There is very considerable progress in the foundation of EGEE on existing networks and on (forthcoming) security architectures. Perhaps the relations to the national CERTs still have to be defined. They could take over a part of the incident processing.

Reply: In EGEE-II a new *vulnerability group* will be established involving a number of related Grid projects to proactively address vulnerability issues identified by the Grid operations groups and user communities. This new group will also liaise with the national Computer Emergency Teams (CERTs).

25. There seems to be a certain danger of functional distribution within the project. Very similar security aspects are worked on in different subprojects.

Reply: In order to better orchestrate all security efforts in the project, for EGEE-II, the different security related groups (JSPG, MWSG, links to other projects security effort) will be coordinated by Security Coordination Group (SCG), under JRA2, and a new group called "Vulnerability Group" involving a number of related Grid projects to proactively address vulnerability issues identified by the Grid operations groups and user communities.

26. EGEE needs to be careful to not let the 6 month deadline drive software choices; EGEE is a long term project which extends beyond the two years, and decisions need to be made with the longer view in mind.

Reply: Following the dissolution of the PTF, the project has decided to create a Technical Coordination Group, including dedicated manpower from all technical activities and led by the

new Technical Director. As many new mechanisms devised for EGEE-II, we are planning to try out this new group before the end of EGEE, in order to make sure it is fully operational by the time EGEE-II starts. One of the goals of this new group is to plan ahead the direction the EGEE middleware needs to take, according to inputs from all stakeholders.

27. Need to raise the level of requiring and disseminating metrics to all groups and collaborators. Part of this is to help build a cache of successes and anecdotes which can be used for "marketing and PR"; part is also to help manage expectations. This becomes even more important as interest and reviews by the EU will continue to increase, especially since they have indicated interest in learning more about how EGEE will impact society and the citizen. In brief, EGEE needs to take the responsibility to proactively manage expectations and market the project as it is highly visible and program officers and EU members already have their ideas of what EGEE will accomplish.

Reply: We have improved our process for press releases and dissemination. For example, every time a new application gets deployed on the Production Service, a press release is sent, in collaboration with the respective application community. This also allows us to generate more PR material around these events, with flyers, high-level description of the application and reference to it on our public website. The release procedure of press releases is also defined such that all NA2 and other partners can piggy-back their own local coverage and material to the main press release, thus improving the impact the project has locally as well. We follow a similar procedure for when important milestones on the Production Service are reached, such as number of site and available storage.

28. Lastly, EGEE tends to focus too much on compute cycles and CPUs as the metric of success; while important, long term success also depends on other resources (data, viz, VOs, etc), people, institutions and collaboration.

Reply: As part of the definition of EGEE-II, we have revisited the *project-level metrics* the project will track and report. We have taken the opportunity to re-focus our metrics on quality of service, instead of volume, which reflect the needs of all users of the Grid.