Initial Proposal for a new EGEE Member

Please take the time to fill this questionnaire in a fully as possible, and send it to Roberto.Barbera@ct.infn.it. Please save a copy of this file and add your responses in the boxes indicated, then save the file and send it to the email address above.

Your Name : P.G.Pelfer	Date: 12/06/05
Organisation :	Your email address *(if different from the one used to send this file)*:
University of Florence and INFN	pelfer@fi.infn.it

Section A: Describe your community

1) Would you describe your community as 'Industrial' or 'Scientific'?
scientific

2) Briefly, how do you feel that your community would benefit from gaining access to the EGEE grid infrastructure? What would you use the Grid's resources for, that could not be done at a local scale?
Modern archaeology, between the historical, anthropological and social sciences, is the more suitable and mature for the application of the Grid technologies. In fact, archaeology is a multidisciplinary historical science, using data and analysis methods from earth sciences, from biology and natural sciences, from physics and chemistry as well as from anthropological and social sciences. From many years, in archaeology, large use is done of computers and complex digital technologies for data acquisition and storage, for quantitative and qualitative data analysis, for data and analysis results visualisation, for mathematical modelling and simulation. The web also is intensively used for results exchange and communication. In recent years is emerging an archaeology, extended at very large region and long time interval, for studying diachronic historical and social process, including the origin of the city and the early state formation. In this cases the archaeological data, produced by excavation and field survey or retrieved from different types of available archives, are not only huge in quantity but also in diversity and complexity, and the computing power needed for their analysis, simulation and visualisation is very large. In fact any material remains, artefacts and ecofacts, macro and microscopic, present on the earth surface, representing the material culture of the past societies is relevant for the archaeology, independently from its esthetical or economical value. Remains should be described according to their basic properties (shape, size, texture, composition, spatial and temporal location), which implies the use of sophisticated procedures for its computer representation: 3D geometry and realistic rendering, among them. Furthermore, data should be related spatially and

(continue)

temporally in complex ways. But archaeology is also a computer intensive discipline. An archaeological site should be understood as a complex sequence of finite states of a temporal trajectory, where an original entity (ground surface) is modified successively, by accumulating things on it, by deforming a previous accumulation or by direct physical modification (building, excavation).

To adequately represent this spatio-temporal trajectory, it must consider a semi-infinite continuum made up of discrete, irregular, discontinuous geometrical shapes (surfaces, volumes) defined by additional characteristics which in turn influence the variation of every archaeological feature. The idea is that interfacial boundaries represent successive phases, and are dynamically constructed. Within them, there should be some statistical relationship between the difference in value of the dependent regionalized variable which defines the discontinuity at any pair of points and their distance apart.

Landscape archaeology, from the other side, reconstruct the evolution of settlement organization on the studied region with a low or high spatio-temporal resolution in relation with the analysed level, intersite, infrasite or regional. For such a precise reconstruction of geomorphology, hydrology, climate, landcover and landuse of the region, based on known data, must be done using models and simulation. Other models relevant to the study of dynamics of ancient societies include demographic models, settlement and urban dynamics and production and exchange models. Model building is time consuming and resource intensive: modeling as a big science. Therefore it need to be able to share models and built from existing models in other branches of science and this can be achieved by the ArchaeoGRID.

The archaeological data are hug amount of unique data and it need care to preserve the physical objects or, at least, the information about them. But also the simulated data must be preserved for a long time because they represent the status of the data interpretation at some date and will be useful for future analysis.("Crisis of Curation")

The structure of the ArchaeoGRID seem to be very complex and cannot include all the needed information and methods of analysis from the other scientific field. In such cases and for specific aspects it seems the best and more convenient way to be interfaced with existing GRID and use database and analysis software available there. As an example, the Digital Elevation Model and the high precision satellite image of the studied region can be acquired interfacing the ArchaeoGRID with the Earth Observation Grid from European Space Agency.

Given the intrinsic nature of archaeological field work, the communication and the information exchange between groups on site and groups working in distant laboratories, museums and universities need fast and efficient communication ways. Telearchaeology lies at the real nature of archaeological endeavor and could be very useful also for education and for diffusion of the archaeological knowledge. A multicast architecture similar to Access Grid for advanced videoconferencing specially tailored for large scale persistent collaboration can be used.

3) How many different countries does your community operate in?
Less than 5?
Between 5 and 15? *between 5 and 15*
More than 15?

4) Within your community, what is your best estimate of how many individuals would be using the Grid infrastructure over the next…	
…12 Months	*…24 Months*
Less than 10? *less than 10*	Less than 10?
Between 10 and 100?	Between 10 and 100? *between 10 and 100*
More than 100?	More than 100?

5) What is your best estimate of the number of sites from your community would be connecting to the Grid infrastructure over the next…	
…12 Months	*…24 Months*
Less than 10? *less than 10*	Less than 10?
Between 10 and 50?	Between 10 and 50? *between 10 and 50*
More than 50?	More than 50?

6) Does your community have particular Security requirements? If so, please briefly describe them.
First of all, archaeological data from excavation and field survey, but also from simulation, if not published in digital form in internet or in paper review and in books, have some private or public owner. In the second place the metadata production, based on agreed semantics and ontologies, is complex and central problem in ArchaeoGRID. High quality metadata will aid the advance in knowledge unification and integration through the use of the GRID information technologies. *Security requirements at data level can be realized by some limitation to data access, respecting the data policy, by an access control list at file level. Metadata access must be leaved for altering and inserting metadata only to metadata producers; the users could have access only to read metadata. At the users level it must me considered the possibility to create groups with defined role for their members, with individual and secure possibility to sign on. Users can be a member of different groups and can have different roles within a group.*

7) Have any members of your community already attended any grid demos or tutorials?
NOT YET
8) If Yes to Q7, approximately how many people attended each event ?
9) If Yes to Q7, how many events of each kind (demos, tutorials etc)?

Section B: Describe your first application

10) Is your first application Industrial or Scientific? Please provide us with a brief description of the application's goals and algorithms
Scientific *Management and analysis of data from excavation and field survey in very large region: Archaeology at the Tops of the World.* *This is an example of how to use telearchaeology to distribute computer resources between scientists studying human settlement in artic and sub-antartic areas. Archaeology in this regions is a hard task, because of climatic and landscape conditions, and also because of the lack of an adequate infrastructure. Additionally, Cultural Heritage in artic and subantartic areas is in real danger of destruction because of climatic change and landscape degradation.* *The problem of the reconstruction of spatio-temporal trajectory - a multidimensional inverse engineering problem - is too big to be adequately computed by usual workstations. Let it see some figures as an example. A normal archaeological site can be of 500 m^2 . To represent it as a space and time shape constrained deformable volume, it need a minimum of 1.000.000 voxels, even in the case of a low resolution representation. Furthermore, the model is multidimensional in nature, what increases the complexity and the computer resources needed. And these figures are only for a single site. A typical area of study is of 10000 sq. km. where it can relate different sites (1 site per 10 sq. km) for a time interval of 6000 years. And all voxels within a computer model of this region should be georeferenced and described with archaeological, geological, environmental and geographical information. The resulting model should contains petabytes of data!* *An additional problem is that computer resources necessary for archaeological research cannot be centralized, given the very nature of archaeological field work. Telearchaeology lies at the very nature of archaeological endeavor.* *As a consequence, it is easy to see that not only scientist, but European society as a whole needs impressive computer resources to generate, curate, analyse and publish historical-archaeological information. Prospective applications go from field surveys, excavations, laboratories to simulations and virtual tools for education and cultural management.* *Only by using Grid-enabled technologies it will be possible to facilitate collaboration with information and resource sharing to all social agents involved in cultural tasks, with security, reliability, accountability, manageability and agility.*

11) Is your first application
CPU Intensive?
Data Intensive?
A combination of both? *both*

12) Is your first application mainly intended for
Interactive use?
Batch use?
A combination of both? *both*

13) Quantitatively evaluate your application in terms of (first application)	
CPU power (SpecInt2000/SpecFp2000 per second per job)	*4 (estimated)*
Memory consumption per job (Megabytes)	*4000 (distributed)*
Disk storage needs per job (Terabytes)	*0.1 (estimated)*
Tape storage needs per job (Terabytes)	*0.2 (estimated)*
Number of jobs per user per year	*20*
Number of users per year	*80*
Network bandwidth requirements (Megabit/sec).	*2 for end user or more*

14) Quantitatively evaluate which percentage of the data of your first application needs to be replicated in more than one site and the average number of copies per elementary replicated data set (e.g., file).
The archaeological data from excavation and field survey are hug amount of unique data and it need care to preserve them for ever. But also the simulated data must be preserved for a long time interval because they represent the status of the data interpretation at some date and will be useful for future analysis. Preservation will be assured if data will be replicated in the larger number of site possible, compatibly with the available resources with the data accessibility.

15) Does your first application have a graphic or a command-line user interface?
both

16) Can your first application be accessed and controlled using a web browser?
yes

17) Is your first application already interfaced to any grid middleware?
NO
18) If yes, to which middleware and in the context of which project?
19) If yes to question 17, can you cite some references (fewer than 10) to related work?

20) Does your first application need third party commercial software to run?
NO
21) If yes, which one(s)?
22) Describe the license under which your first application can be distributed on the grid infrastructure.
23) Describe concisely the security requirements of your first application.
look at the question 6

24) If you answered 'Yes' to questions 17 and 20, describe concisely (max 500 words) what problems you foresee to successfully port your first application on a distributed computing environment like the European Grid infrastructure.

Section C: Describe your second application

25) Is your second application *Industrial* or *Scientific*? Please provide us with a brief description of the application's goals and algorithms
Scientific *Reconstruction of a central historical process connected to a very large region based on existing archaeological data and simulated ones: the Origins of the City and of the State in the Mediterranean area between XI and VIII centuries B.C.* *This is one of the most important research subjects for understanding modern identity of Europe. It is the case of the study of an historical process related with a large region, the Mediterranean area, and with a not too long time interval. In fact the emergence of urban centers along the Mediterranean costal lines take a short period equivalent to few generations. Existing data in such case can be very large amount, not homogeneous, from many different sources and simulated data could be also larger. Their production and management, their organization and archiving in large and accessible database are problems that cannot find solution in the traditional approach, but require new solution by using Grid-enabled technologies* *However, the huge quantity and diversity of necessary data, the lack of a central repository for environmental and social information and the complex simulation tools needed to understand social processes imply the use of Grid technologies. It is essential to start combining, integrating, comparing and contrasting data collected by individual archaeological projects. Certain (especially demographic and economic) questions of supra-regional scope and significance can no longer be reliably addressed without drawing upon collated archaeological survey-data. It need to bring disparate data-sets together in a GIS databank for Mediterranean archaeology for compare, statistically, the relative developments of Iron Age settlement patterns in Turkey, Greece, Italy, France, Spain, Libya and Tunisia, towards analysing if there is, or not, an impact of colonization during that period and his rate and nature. The distribution of tool types and materials, or pottery wares and forms, from surveys in various periods could refine our knowledge of trade and technology patterns across the Mediterranean. The impact of human settlement and resource extraction upon the environment could be approached, towards solving the problem of deforestation and erosion in the ancient world. Finally, historians could add such collected survey data to the other sets of information they use to build large-scale, long-term explanations of Mediterranean polities and societies. In this perspective it will need to design and implement a simulation software framework to examine the dynamic social process of the urban process. The dynamic software model that concurrently represent both natural and social processes must have levels of fidelity and granularity that make feasible to represent their interactions over a large range of spatial and temporal scale.* *Model building is time consuming and resource intensive: modeling as a big science. Simulation in archaeology can be considered like an heuristic device to start the analysis of large and complex problems as the origin of the city.*

26) Is your second application
CPU Intensive?
Data Intensive?
A combination of both? *both*

27) Is your second application mainly intended for
Interactive use?
Batch use?
A combination of both? *both*

28) Quantitatively evaluate your second application in terms of	
CPU power (SpecInt2000/SpecFp2000 per second per job)	*4 (estimated)*
Memory consumption per job (Megabytes)	*4000 (distributed)*
Disk storage needs per job (Terabytes)	*0.1 (estimated)*
Tape storage needs per job (Terabytes)	*0.2 (estimated)*
Number of jobs per user per year	*20*
Number of users per year	*80*
Network bandwidth requirements (Megabit/sec).	*2 or more*

29) Quantitatively evaluate which percentage of the data of your second application needs to be replicated in more than one site and the average number of copies per elementary replicated data set (e.g., file).
The archaeological data from excavation and field survey are hug amount of unique data and it need care to preserve them for ever. But also the simulated data must be preserved for a long time interval because they represent the status of the data interpretation at some date and will be useful for future analysis. Preservation will be assured if data will be replicated in the largest possible sites number, compatibly with the available resources and with data accessibility.

30) Does your second application have a graphic or a command-line user interface?
both

31) Can your second application be accessed and controlled using a web browser?
yes

32) Is your application no. 1 already interfaced to any grid middleware?
NO
33) If yes, to which middleware and in the context of which project?
34) If yes to question 17, can you cite some references (less than 10) to related work?

35) Does your application no. 1 need third party commercial software to run ?
NO
36) If yes, which one(s)?
37) Describe the license under which your second application can be distributed on the grid infrastructure.
38) Describe concisely the security requirements of your second application.
look at the question 6

39) If yes to questions 32 and 38, describe concisely what problems you foresee in successfully porting your second application to a distributed computing environment like the European Grid infrastructure?

Section D: Describe your commitment

40) How many people / FTE's can your community dedicate over the next 12 months to port your application(s) to the grid infrastructure in collaboration with grid experts?
three or more people

41) From how many sites and with what distribution among sites?
one people from three or more sites

42) How many people/FTE's can your community dedicate in the next two years to port your application(s) to the grid infrastructure in collaboration with grid experts?
six or more people

43) How much and of what nature of computing and/or storage resources can your community introduce to the grid infrastructure to run your application(s) in the next 12 months? From how many sites and with what distribution among sites?
local storage and computing resource already available by EDG and by PC farm and (super)computing centres from three sites.

44) How much and of what nature of computing and/or storage resources can your community introduce to the grid infrastructure to run your application(s) in the next 2 years? From how many sites and with what distribution among sites?
some PC based computing and storage resources

45) Would your resources be made available to the EGEE Grid for use by other community members? If YES, which is the percentage of use that can be allowed to other communities for their applications?
yes; 30%

46) How many people / FTE's can your community dedicate over the next 12 months to install, upgrade and manage grid middleware the community's Sites? From how many Sites and with what distribution among them?
two

47) How many people / FTE's can your community dedicate over the next 2 years to install, upgrade and manage grid middleware on the community's Sites? From how many Sites and with what distribution among them?
two

Feedback

As a potential user of the EGEE Grid, you will be expanding the Grid simply by being a part of it. We are keen to expand the Grid as much as possible and to do this we need to get EGEE promoted in the right places, so the right people get to know of it. You can help us by answering the following questions relating to the EGEE Project's public profile.

How did you first hear about EGEE?
Website? *CERN website*
Word-of-Mouth?
Publication?
Event? *IST conferences*
Other? *CERN laboratory*

In your particular field, at which event(s) would you most expect to see EGEE Grid activities publicised or promoted?
EGEE conferences, Computer Application in Archaeology conferences, IEEE software conference, workshops and event organized by ArchaeoGRID groups.

In your particular field, in which publication(s) would you most expect to see EGEE Grid activities publicised?
International Reviews in Archaeology and Anthropology, International Reviews in Complex Social Systems Analysis, IEEE Software Reviews

Which publication(s) do you read on a regular basis? (include newspapers, periodicals and any other)
International Reviews in Archaeology, Anthropology and Ancient History; International Reviews on Computer Application in Archaeology; Geographical and GIS Reviews; Reviews on Applied Statistics; Reviews in Agriculture and Technology History; Complex Systems Analysis Reviews.

Thank you for your time.