



Key Directions for Research Libraries e-Infrastructure and the HEP community

Federico Ruggieri, ane Hunter, University of Queensland

Background

The HEP community, became involved in Grid computing in the late nineties to solve the huge LHC computational problem which was starting to be investigated (after an initial under-evaluation). At that time, client-server and meta-computing were the frontier and first implementations of Computer Farms were appearing. The largest problem anyway was the huge amount of data expected to be produced and analyzed (tens of PetaBytes per year).

The 'social' challenge was to allow thousands of physicists to access those data easily many countries in different continents. It was also clear that even taking into account the Moore's Law for the Computing power evolution, the CERN budget traditionally dedicated to Computing resources was largely insufficient. There was no obvious solution on the market and such a worldwide enterprise requested new approaches. Several 'new' technologies were proposed like Object Oriented (OO) Programming and OO DataBases, and several Research and Development projects proposed to solve it.

High Energy Physics (HEP) computing, on the other hand, is a typical **High Throughput Computing** that allows a very simple or 'natural' parallelization based on the replica of the application program and the Event based data structure.

The basic approach proposed was to distribute the load of LHC computing in the various laboratories with [CERN](#) being the Data Source and the main repository of such data. The model, proposed by the [MONARC Project](#) defined a few levels or 'tiers', with CERN as a Tier0 and the other Regional Centres as Tier1s with Tier2s underneath.

Grid Computing appeared as a natural answer to those requirements.

HEP - the Grid and e-Infrastructure

The huge amount of data to be processed and analysed is anyway the real big challenge of the LHC Computing Grid and the usage of catalogs, replicas and tag DataBases is the present way of addressing it. The access of those data will be restricted to the researchers involved (thousands) during the whole duration of the experiments (ten years at least) for obvious reasons of competitiveness among them.



In the last few years, the scenario of **international collaboration** in Research and beyond has swiftly evolved with the gradual but impressive deployment of large bandwidth networks. A number of advanced services and applications have been using these networks, enabling new ways of remote collaboration. The environment resulting from the integration of networking and other resources, such as **Digital Libraries, computing, storage, instruments** and related systems is also known as **e-Infrastructure**.

eInfrastructure - Bridging the gap for the developing world

In the most advanced economies, knowledge is nowadays one of the major elements of progress and economic welfare and e-Infrastructures are, in turn, one of the major enablers of development in a knowledge economy. On the other hand, this threatens to widen the digital gap between developing economies and the most advanced ones, where knowledge is a commodity and an important share of the budget of companies and governments is allocated on R&D and on Education: the latter gets, as a return of their conspicuous investments, more and more advanced infrastructures and techniques that enable in turn new developments, while the former, taken off late and with less resources and urged by more fundamental needs, seem incapable to reduce the Gap.

At a first glance, developing countries have much more compelling priorities to fund than building e-Infrastructures. Nevertheless, it is important to understand the role of e-Infrastructures in breaking the loop of mere subsistence. Science is at the basis of long-term innovation in production activities, and **digital infrastructures** are in turn necessary to allow researchers to participate to frontier scientific activities, to share competences and experiences with their counterparts all around the world, thus being up with the most recent tools and methods. On the other hand, there are several examples of points of excellence in the Mediterranean area such as [Bibliotheca Alexandrina](#) in Egypt. Moreover a few initiatives tend to deploy new important research infrastructures (e.g. [SESAME](#) in Jordan) which will produce relevant amounts of data and knowledge that need to be shared. The growth of digital data and libraries in the African and Middle-East countries will provide a large amount of unique knowledge related to different cultures, languages and histories.

Moving Forward

The vision for the future should include a few cornerstones on top of which a consistent architecture should be built:

- New hardware and software technology will naturally be helpful but **standardisation** of solutions will need a reasonable dedicated effort. The aim will not be to create standards, but to agree on a series of good compromises.



- **Policy** of management to maintain in the long term the important knowledge that has been accumulated. This will include the policy by which the knowledge to be kept will be distinguished by the data that can be discarded.
- **Ubiquitous access** to libraries and knowledge is a must in a society that is already engaged in high mobility and world wide connectivity. New service for the general public can be envisaged and the network connectivity will no more be the limiting factor.
- **Virtual Communities and commonalities** will be relevant in a globalised world where researchers, groups of interest and even collaborating companies require personalised access to huge amounts of knowledge and data. Related to this will be the relevance of an easy access to high level services that will be able to adapt themselves to the different needs of the customers.

A number of steps need to be taken if this vision is to be achieved. The creation of stable **interoperable e-Infrastructures** with similar services available in Europe and in other parts of the world, is of utmost importance. Dedicated projects that will explore the field and suggest possible ways of creating an adaptable architecture for library based high level services. Inclusion of communities in Africa, Middle-East and Asia which can bring a different cultural approach and innovating ideas.

About the Author

Dr. Federico Ruggieri, is an INFN senior physicist and Director of Research in the INFN Section of Roma Tre. He has spent most of his professional life working on On-Line and Off-line Computing Systems for High Energy Physics experiments at CERN (The European Particle Physics Laboratory) and at Frascati, INFN National Laboratory.



His experience includes working on various projects including the Condor Project, DATA Grid and is presently the Project Manager for EUChinaGRID and EUMEDGRID. He has also been Director at the INFN National centre responsible for Research and Development of Informatics and Telematics Technologies (CNAF), and been involved in the development of the Networks for Research in Italy like INFNet the INFN Wide area network, and GARR the National Research and Academic Network. From 2000 to 2006 he was member of the Scientific and Technical Advisory Committee of GARR. Since 2004 he has been a member of the Scientific and Technical Committee of CASPUR a Consortium of Universities and a Computing Centre for High Performance Computing. He is also a Qualified Expert in Informatics and Telematics for the Italian Ministry of University and Research (MUR).