From eScience to eBusiness

The Industry and Research Forum

As part of its dissemination activity a network of active and interested institutes have been brought together to form the DataGrid Industry and Research Forum. The Forum brings together researchers and system developers in different disciplines (Earth Observation, Biology, Physics and Computer Science) from tens of countries. The list of current participants includes many research centres, universities and industries and is open to new participants with the aim of creating the critical mass of experience and interest for the deployment of the Grid technologies. The Industry & Research Forum will be supported for the lifetime of the project, with activities focused on the wide dissemination of the results of the DataGrid project within the industrial and scientific communities and with the aim of stimulating industry early adopters to join the first movers in Grid computing. The Forum is the main exchange place for information, dissemination and potential exploitation of the Data Grid results. For such a purpose, a special interest mail-list is open to all interested parties. This list is used for an open discussion of the issues and announcement of results as they become available. The Forum also allows unfunded participation from countries and parties which otherwise would have difficulty joining an EU consortium.

Principal Contractors

European Organisation
for Nuclear Research
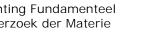
Centre National de la
Recherque Scientifique

European Space Agency

Istituto Nazionale di Fisica Nucleare

Particle Physics and Astronomy
Research Council

Stichting Fundamenteel
Onderzoek der Materie

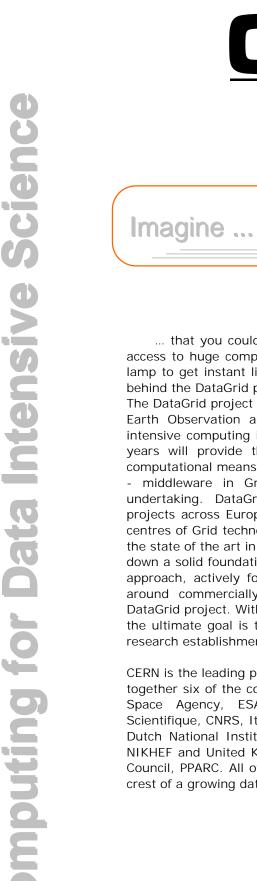## Assistant Contractors

– CESNET - Zaymove Sdruzeni Pravnickych Osob
– Consiglio Nazionale delle Ricerche (CNR)
– Helsinki Institute of Physics
– Institut de Fisica d'Altes Energies (IFAE)
– Konrad-Zuse-Zentrum für Informationstechnik
– Ruprecht-Karls-Universität Heidelberg
– Swedish Natural Science Research Council (NFR)
– Commissariat à l'energie atomique (CEA)
– Computer and Automation Research Institute,
   Hungarian Academy of Sciences (MTA SZTAKI)
– Istituto Trentino di Cultura (IRST)
– Royal Netherlands Meteorological Institute (KNMI)
– Stichting Academisch Rekencentrum Amsterdam (SARA)
- Communication & Systèmes (CS)
Ξ DATAMAT Ingegneria dei Sistemi S.p.A
- IBM United Kingdom Limited

Industry and Research Forum
INFN-Sezione di Padova, Italy
Via Marzolo, 8 35131 Padova
Contact Name:
Mirco Mazzucato
e-mail: mazzucato@padova.infn.it

Project Management Office
CERN - Switzerland
1211 Geneva 23
Project Manager:
Fabrizio Gagliardi CERN-IT
Project Secretariat:
e-mail: fabrizio.gagliardi@cern.ch
Tel: +41.22.7672374

Project Dissemination Office
CNR – CED Italy
P.le Aldo Moro, 7 00185 Rome
Contact Name:
Ms. Susanna Tosi
e-mail:s.tosi@cedrc.cnr.it
Tel: +39.06.44933205

Data Grid
Computing for Data Intensive Science

DataGRID

Imagine ...

... that you could plug your computer into the wall and have direct access to huge computing resources immediately, just as you plug-in a lamp to get instant light. Far from being science fiction, this is the idea behind the DataGrid project and what it will lead to in the future.

The DataGrid project brings together researchers from Biological Science, Earth Observation and High-Energy Physics where large-scale, data-intensive computing is essential. The needs of these fields over coming years will provide the data - the DataGrid project will provide the computational means to handle them. Developing the necessary software - middleware in Grid parlance - to meet this vision is a global undertaking. DataGrid complements and works with national Grid projects across Europe and benefits from collaborating with the leading centres of Grid technology around the world. In doing so, it is extending the state of the art in large-scale, data-intensive computing, and is laying down a solid foundation for European industry to build on. The European approach, actively fostered by the European Union, is to build a Grid around commercially available products using open software – the DataGrid project. With the conceptual idea of creating a World Wide Grid, the ultimate goal is to provide computing on-tap, not only for scientific research establishments, but also for libraries, schools and homes.

CERN is the leading partner in the EU-funded DataGrid project that brings together six of the continent's leading research agencies - the European Space Agency, ESA, France's Centre National de la Recherche Scientifique, CNRS, Italy's Istituto Nazionale di Fisica Nucleare, INFN, the Dutch National Institute for Nuclear Physics and High Energy Physics, NIKHEF and United Kingdom's Particle Physics and Astronomy Research Council, PPARC. All of these bodies are working in fields that are on the crest of a growing data-wave.

CERN, Weighty Questions

The European Organisation for Nuclear Research

, CERN, is building a particle accelerator more powerful than any before. Called the Large Hadron Collider, LHC, it will address some very weighty questions about the fundamental particles of matter and the forces that act between them. Why, for example, are some particles so much heavier than others? And why do particles have mass at all? The answers may be close at hand in the form of an all pervading presence called the Higgs field. Each particle's mass could be determined by the strength of its interaction with this field. Evidence for the Higgs field will be so rare as to make finding the proverbial needle in a haystack seem like child's play. A needle in 20 million haystacks would be closer to the mark. For this reason, the LHC will provide data at a tremendous rate - equivalent to every person on the planet talking into 20 telephones at once. Even after filtering, the data that will have to be stored and analysed in institutes around the world each year is equivalent to a pile of CD-ROMs standing about a kilometre high. The Grid provides the solution, allowing LHC scientists to exploit major distributed resources, such as processing power, and databases on a much larger scale, more effectively, and far more transparently than ever before. The model that has been adopted for LHC computing is multi-tiered, with a large centre at CERN linked by a high-speed network to approximately ten in national or regional data centres, each of which will have significant resources. These in turn fan out to local regional centres with the ability to perform smaller subsets of analysis and so on to the individual universities and institutes that make up the global LHC community.

ESA, Earth Watching

The European Space Agency

manages several Earth Observation satellite missions. These involve the downloading of about 100 Gigabytes per day of images and this number will grow in the near future by a factor of about five, with the launch of the ENVISAT satellite due in 2001. Dedicated ground infrastructures have been set up to handle the data produced by instruments onboard the satellite. Some 800 Terabytes are already stored in mass storage facilities, and this is set to grow in the future. The Earth Science community will benefit greatly from the DataGrid project. The Grid will guarantee an improved way to access large volumes of data stored in a distributed European-wide archive, and provide far greater computing muscle. The new Grid architecture will also strengthen ongoing large distributed data access and processing activities. As the Grid distributed data archive access approach fits the requirements of the Earth Science user community, a number of testbed prototyping activities are planned to validate the platform concept. A complete application that involves the use of atmospheric ozone data has been selected as a specific testbed. This will constitute the basis for the potential evolution of the Grid platform towards much larger computational and data processing requirements.

CNRS, Exploring the Genome

The human genome

, with its 3.5 billion pairs of chemicals that make up the DNA in every human cell, is in the media spotlight, but it is just one of many genomes that have been sequenced. Annotation of genomes requires larger and larger databases, and the integration of a testbed dedicated to biology into the DataGrid project opens up new perspectives for computing resources and data storage. The storage and exploitation of genomes and of the huge flux of data coming from post-genomics puts growing pressure on computing tools - such as databases and code management - and on computing resources within laboratories. The emerging computational Grid technologies offer a solution. The DataGrid project's biology testbed will provide a platform for new algorithms for data mining, new databases, code management, graphical interface tools, and new strategies for in silico experimentation. Grid technology will be deployed to optimize the use of computing resources for these highly intensive applications, and will facilitate sharing of genomic databases for the benefit of international cooperation. Another area of application in biology is the processing of medical images. From image acquisition and processing to interpretation and archiving, many distributed Grid resources have to be allocated to handle and process very large digital images in quasi real time. The functionality of the Grid will facilitate and enhance cooperation between distributed experts allowing them to work on common objects irrespectively of the location of the resources they will use.