The EU DataGrid

The European DataGrid Project Team

http://www.eu-datagrid.org/

Tutorial Roadmap

- Introduction
- Security
- The Testbed
- Middleware Issues
  - Job Submissions
  - Data Management
  - Information and Monitoring
  - Fabric Management
- Applications
  - HEP
  - EO
  - Biology

The EU DataGrid: Introduction

The European DataGrid Project Team

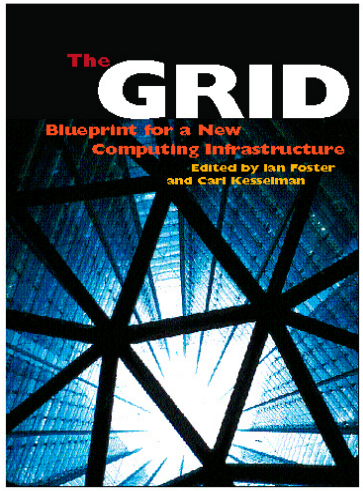
http://www.eu-datagrid.org/

Contents

- European DataGrid (EDG) Project scope
- EDG structure
- Major components of EDG Middleware
- DataGrid in Numbers
- Relation to Sister Projects

The Grid Vision

- Flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resource
  - From "The Anatomy of the Grid: Enabling Scalable Virtual Organizations"
- Enable communities ("virtual organizations") to share geographically distributed resources as they pursue common goals -- *assuming the absence* of…
  - central location,
  - central control,
  - omniscience,
  - existing trust relationships.



Grids: Elements of the Problem

- Resource sharing
  - Computers
  - Storage
- Security
  - Resource sharing always conditional
  - issues of trust, policy, payment, …
- Coordinated problem solving
  - Beyond client-server
  - Distributed data analysis
- Virtual Organisations
  - Community overlays on classic organisational structures
  - Large or small, static or dynamic

EU DataGrid

- DataGrid funded by European Union whose objective to exploit and build the next generation computing infrastructure providing intensive computation and analysis of shared large-scale databases.
- Enable data intensive sciences by providing world wide Grid test beds to large distributed scientific organizations ("Virtual Organizations, VOs")
- Duration: Jan 1, 2001 - Dec 31, 2003
- Applications/End Users Communities: HEP, Earth Observation, Biology
- Specific Project Objectives:
  - Middleware for fabric & grid management
  - Large scale testbed
  - Collaborate and coordinate with other projects
  - Contribute to Open Standards and international bodies

    (GGF, Industry&Research forum)

DataGrid Main Partners

- CERN – International (Switzerland/France)

- CNRS - France

- ESA/ESRIN – International (Italy)

- INFN - Italy

- NIKHEF – The Netherlands

- PPARC - UK

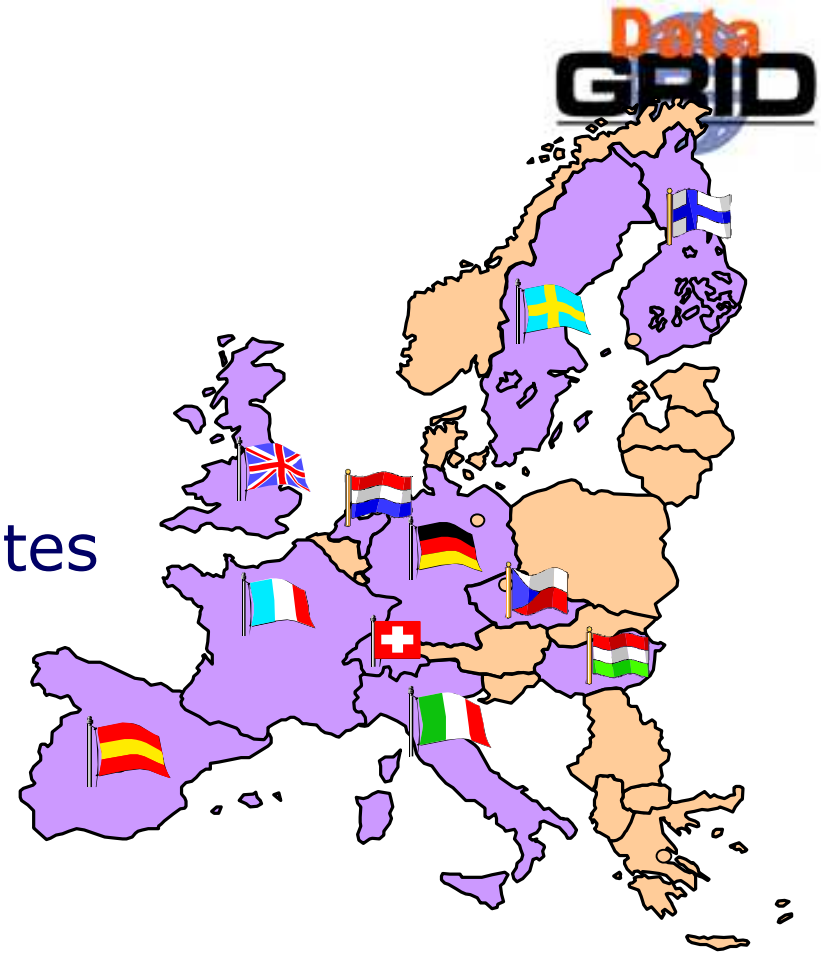
Assistant Partners

Industrial Partners

- Datamat (Italy)
- IBM-UK (UK)
- CS-SI (France)

Research and Academic Institutes

- CESNET (Czech Republic)
- Commissariat à l'énergie atomique (CEA) – France
- Computer and Automation Research Institute, Hungarian Academy of Sciences (MTA SZTAKI)
- Consiglio Nazionale delle Ricerche (Italy)
- Helsinki Institute of Physics – Finland
- Institut de Fisica d'Altes Energies (IFAE) - Spain
- Istituto Trentino di Cultura (IRST) – Italy
- Konrad-Zuse-Zentrum für Informationstechnik Berlin - Germany
- Royal Netherlands Meteorological Institute (KNMI)
- Ruprecht-Karls-Universität Heidelberg - Germany
- Stichting Academisch Rekencentrum Amsterdam (SARA) – Netherlands
- Swedish Research Council - Sweden



Project Schedule

- Project started on 1/Jan/2001
- Series of deployments to the "application testbed"
  - Early 2001
    - Globus 1 only - no EDG middleware
    - To understand some deployment issues
  - Early 2002
    - First release of EU DataGrid software to tolerant users within the project
  - Subsequent releases
    - Adding functionality
    - Concentrating on achieving production quality
    - This is where we are today…
  - Final release late 2003
    - Not much added functionality
- Project stops on 31/Dec/2003

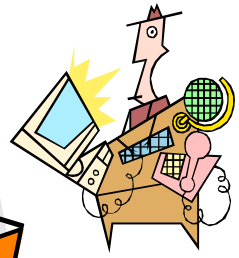
DataGrid Work Packages

- The EDG collaboration is structured in 12 Work Packages
  - WP1: Work Load Management System
  - WP2: Data Management
  - WP3: Grid Information and Monitoring
  - WP4: Fabric Management
  - WP5: Storage Element / Storage Resource Manager
  - WP6: Testbed and demonstrators – *Production quality International Infrastructure*
  - WP7: Network Monitoring
  - WP8: High Energy Physics Applications
  - WP9: Earth Observation
  - WP10: Biology
  - WP11: Dissemination
  - WP12: Management

Bodies

- Project Management Board (PMB)
  - Look after the politics
- Project Technical Board (PTB)
  - Meets infrequently to approve deliverables
  - Work Package Managers meet weekly
- Architecture Task Force (ATF)
  - Define interfaces
- Quality Assurance Group (QAG)
  - Define relevant standards
- Integration Team (ITeam)
  - Glue it all together

EDG Interfaces

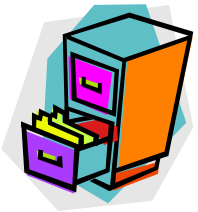


System Managers

Application Developers

Scientists

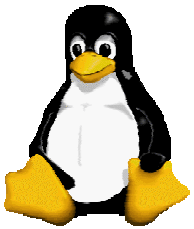
Certificate Authorities



File Systems

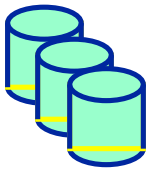
User Accounts

- Next slides show major components of middleware developed by WP1-5 and 7

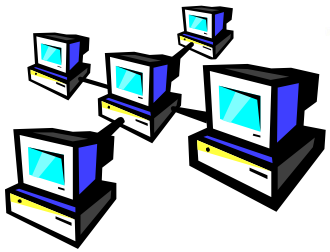


Operating Systems

Mass Storage Systems
HPSS, Castor



Storage Elements



Computing Elements

Batch Systems
PBS, LSF

The EDG WMS

- The user interacts with Grid via a **Workload Management System** (WMS)
- The Goal of WMS is the **distributed scheduling and resource management in a Grid environment.**
- What does it allow Grid users to do?
  - To submit their jobs
  - To execute them on the "best resources"
    - The WMS tries to optimize the usage of resources
  - To get information about their status
  - To retrieve their output

Computing Element

- Is a Grid Job Queue
  - Publishes information about itself
  - Checks the job is permitted
  - Sends it to an an appropriate internal queue

SRM: Storage Resource Manager

- SRM subset implementation
  - A defacto international standard for Storage Resource Management
- Web service uses Java AXIS and EDG security
- Supports multiple VOs
- Functions
  - Writing a file
  - Reading a file

Replica Manager

- Hides the SRM
- Coordinates use of
  - Replica Location Service
  - Replica Metadata Catalog
  - Replica Optimization Service

R-GMA: Information & Monitoring

- Relational implementation of GMA from GGF
- Makes use of GLUE schema
- Interoperable with MDS
- Deals with information on
  - The Grid itself
    - Resources and Services (for which the Globus MDS is a common solution)
    - Job status information
  - Grid applications
    - This is information published by user jobs.

WP6: TestBed Integration

- Exact definition of RPM lists (components) for the various testbed machine profiles (CE, RB, UI,, WN, IC etc.) – check dependencies
- Perform preliminary centrally (CERN) managed tests on EDG m/w before green light for spread EDG testbed sites deployment
- Provide, update end user documentation for installers/site managers, developers and end users
- Define EDG release policies, coordinate the integration team staff with the various WorkPackage managers – keep high inter-coordination.
- Set up the Authorization Working Group to manage authorization policies on the testbed

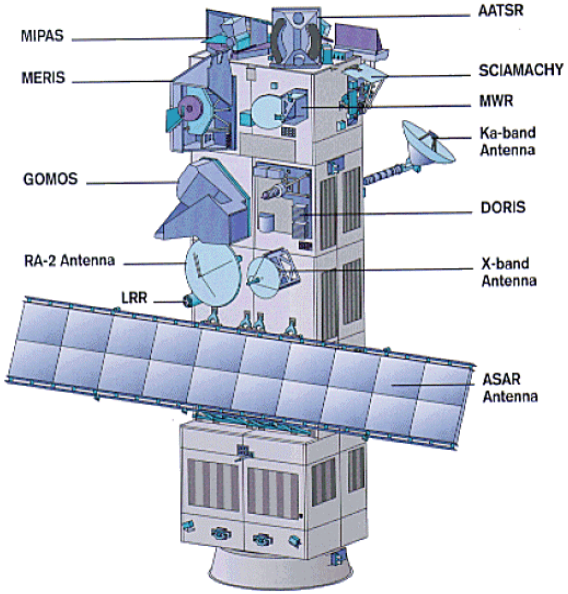
Grid aspects covered by EDG

VOMS	Provides certificate with VOs, groups and roles	**RGMA: Information & Monitoring**	Provides info on resource utilization & performance
User Interface	Submit & monitor jobs, retrieve output	**Grid Fabric Management**	Configure, installs & maintains grid sw packages and environ.
Workload Management System	Manages submission of jobs to Res. Broker, obtains information and retrieves output	**Network performance**	Provides efficient network transport, bandwidth monitoring
Computing Element	Gatekeeper to a grid computing resource	**Testbed admin.**	Certificate auth.,user reg., usage policy etc.
Storage Resource Manager	Grid-aware storage area	**Applications**	HEP, EO, Biology
Replica Manager	Replicates and locates data		

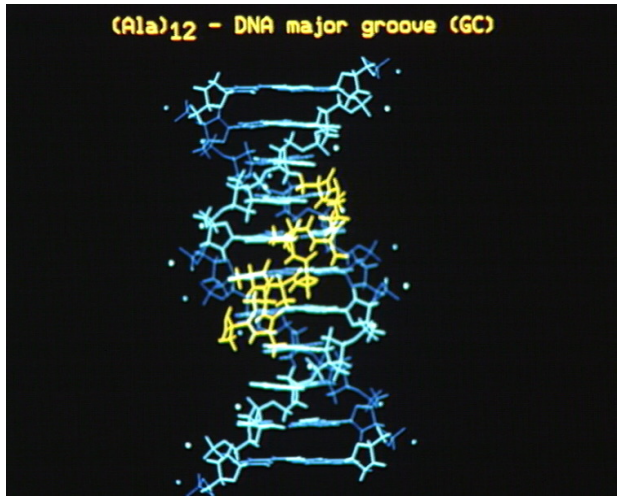
Applications (WP8-10)

High Energy Physics

Earth Observation Science Applications



Biomedical Applications



DataGrid in Numbers (Out of date)

People

>350 registered users

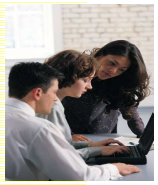
12 Virtual Organisations

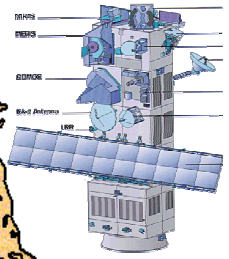
16 Certificate Authorities

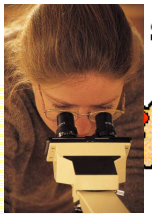
>200 people trained

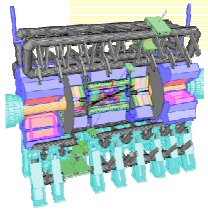
278 man-years of effort

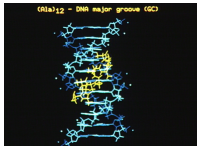
100 years funded

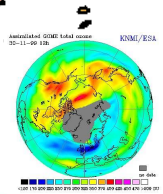












Testbeds

>15 regular sites

>10'000s jobs submitted

>1000 CPUs

>5 TeraBytes disk

3 Mass Storage Systems

Software

50 use cases

18 software releases

>300K lines of code

728454 as measured by SLOCCount on 25 June

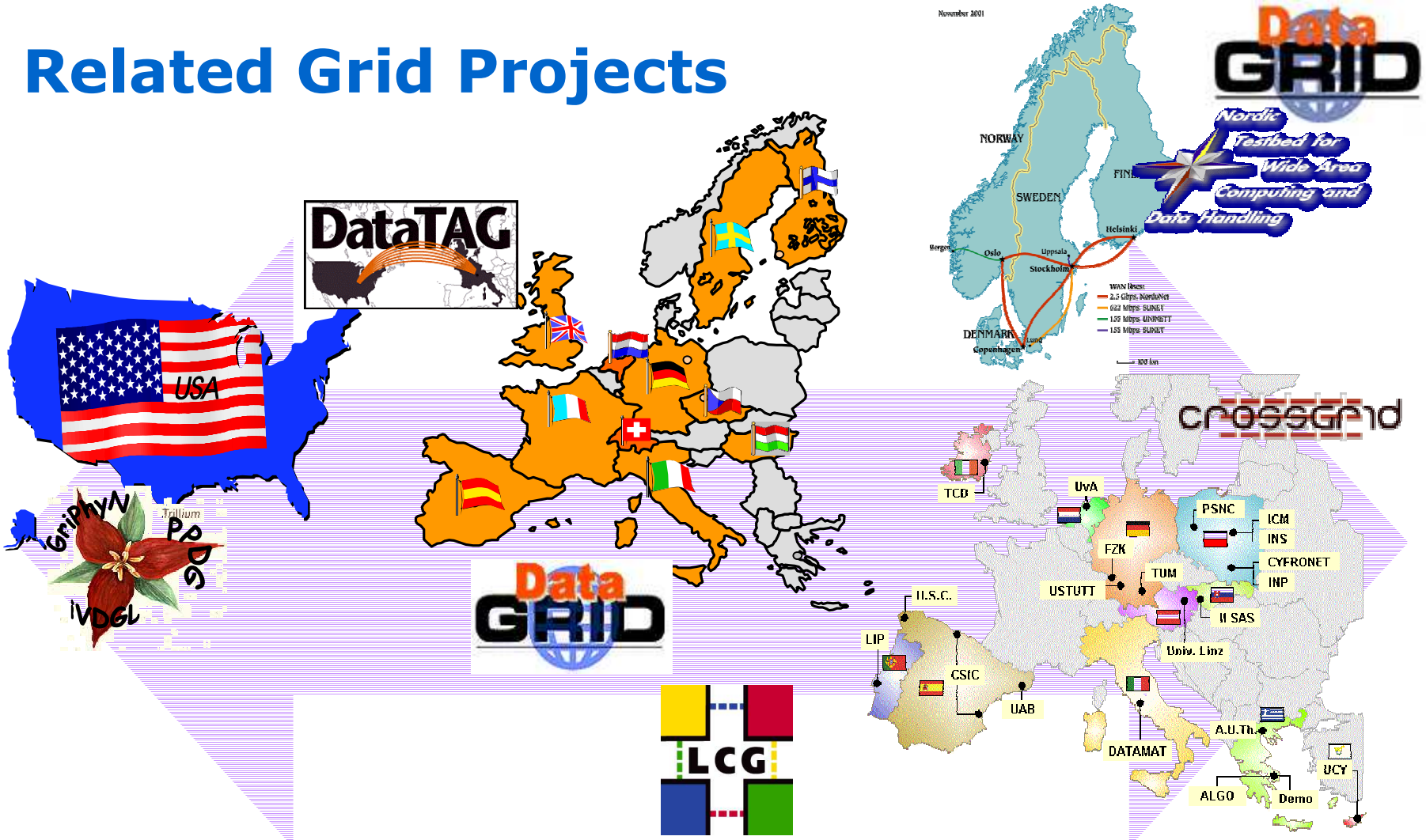
Scientific applications

5 Earth Obs institutes

9 bio-informatics apps

6 HEP experiments

Related Grid Projects



Through links with sister projects, there is the potential for a truly global scientific applications grid
Demonstrated at IST2002 and SC2002 in November