

Datagrid initiatives in Europe

Larry Price

ANL

PPDG Meeting

July 13, 2000

...Thanks to F. Gagliardi and F. Ruggieri

PPDG July 2000

EU DataGrid Initiative

**Proposal Title: "Research and Technology
Development for an International Data Grid"**

**European level coordination of national initiatives &
projects**

Principal goals:

- Middleware for fabric & Grid management
- Large scale testbed - major fraction of one LHC experiment
- Production quality HEP demonstrations
 - "mock data", simulation analysis, current experiments
- Other science demonstrations

Three year phased developments & demos

Complementary to other GRID projects

- **EuroGrid: Uniform access to parallel
supercomputing resources**

PPDG July 2000

L. Price 2

Preliminary programme of work

- WP 1 Grid Workload Management (C. Vistoli/INFN)**
- WP 2 Grid Data Management (B. Segal/CERN)**
- WP 3 Grid Monitoring services (R. Middleton/PPARC)**
- WP 4 Fabric Management (T. Smith/CERN)**
- WP 5 Mass Storage Management (J. Gordon/PPARC)**
- WP 6 Integration Testbed (F. Etienne/CNRS)**
- WP 7 Network Services (C. Michau/CNRS)**
- WP 8 HEP Applications (F. Carminati/CERN)**
- WP 9 EO Science Applications (L. Fusco/ESA)**
- WP 10 Biology Applications (C. Michau/CNRS)**

PPDG July 2000

L. Price 3

List of Principal Contractors

Principal Contractor	Country Code	Abbreviation for Anonymous Part B
CERN (coordinator)	F	CO
CNRS	F	CR2
ESA	I	CR3
INFN	I	CR4
NIKHEF	NL	CR5
PPARC	UK	CR6

PPDG July 2000

L. Price 4

Current list of Assistant Contractors

Assistant Contractor	Country Code	Abreviation for Anonymous Part B	Assistant to Principal
Istituto Trentino Di Cultura	I	AC7	CO
Helsinki Institute of Physics	FIN	AC8	CO
Science Research Council	S	AC9	CO
Zuse Institut Berlin (ZIB)	D	AC10	CO
University of Heidelberg (KIP)	D	AC11	CO
CS Systèmes d'Information	F	AC12	CR2
CEA/DAPNIA	F	AC13	CR2
IFAE	E	AC14	CR2
Datamat	I	AC15	CR4
CNR (Italian Research Council)	I	AC16	CR4
CESNET	CZ	AC17	CR4
Koninklijk Nederlands Meteorologisch Instituut (KMNI)	NL	AC18	CR5
Stichting Academisch Rekencentrum Amsterdam (SARA)	NL	AC19	CR5
SZTAKI	HU	AC20	CR6
IBM	UK	AC21	CR6

PPDG July 2000

L. Price 5

Status

Prototype work already started at CERN and in most of collaborating institutes (Globus initial installation and tests)

Proposal to the EU submitted on May 8th, being reviewed by independent EU experts now

HEP and CERN GRID activity explicitly mentioned by EU official announcements

(http://europa.eu.int/comm/information_society/eeurope/news/index_en.htm)

Project presented at DANTE/Geant, Terena conference, ECFA (tomorrow)

PPDG July 2000

L. Price 6

Near Future Plans

- Quick answers to the referee report (mid July).
- Approval (hopefully) by the EU-IST Committee (12-13 July).
- Technical Annex Preparation (July-August).
- Work Packages workshop in September.
- Contract negotiation with EU (August- October)
- Participation to conferences and workshops (EU Grid workshop in Brussels, iGRID2000 in Japan, Middleware workshop in Amsterdam).

PPDG July 2000

L. Price 7

WP 1 GRID Workload Management

Goal: define and implement a suitable architecture for distributed scheduling and resource management in a GRID environment.

Issues:

- Optimal co-allocation of data, CPU and network for specific "grid/network-aware" jobs
- Distributed scheduling (data and/or code migration) of unscheduled/scheduled jobs
- Uniform interface to various local resource managers
- Priorities, policies on resource (CPU, Data, Network) usage

PPDG July 2000

L. Price 8

WP 2 GRID Data Management

Goal: to specify, develop, integrate and test tools and middle-ware infrastructure to coherently manage and share petabyte-scale information volumes in high-throughput production-quality grid environments

WP 3 GRID Monitoring Services

Goal: to specify, develop, integrate and test tools and infrastructure to enable end-user and administrator access to status and error information in a Grid environment.

Goal: to permit both job performance optimisation as well as allowing for problem tracing, crucial to facilitating high performance Grid computing.

WP 4 Fabric Management

Goal: to facilitate high performance grid computing through effective local site management.

Goal: to permit job performance optimisation and problem tracing.

Goal: using experience of the partners in managing clusters of several hundreds of nodes, this work package will deliver a computing fabric comprised of all the necessary tools to manage a centre providing grid services on clusters of thousands of nodes

PPDG July 2000

L. Price 11

WP 5 Mass Storage Management

Goal: Recognising the use of different existing MSMS by the HEP community, provide extra functionality through common user and data export/import interfaces to all different *existing* local mass storage systems used by the project partners.

Goal: Ease integration of local mass storage system with the GRID data management system by using these interfaces and through relevant information publication.

PPDG July 2000

L. Price 12

WP 6 Integration testbed

Goals:

- plan, organise, and enable testbeds for the end-to-end application experiments, which will demonstrate the effectiveness of the Data Grid in production quality operation over high performance networks.
- integrate successive releases of the software components from each of the development workpackages.
- demonstrate by the end of the project testbeds operating as production facilities for real end-to-end applications over large trans-European and potentially global high performance networks.

PPDG July 2000

L. Price 13

WP 7 Networking Services

Goals:

- review the network service requirements of DataGrid and make detailed plans in collaboration with the European and national actors involved.
- establish and manage the DataGrid VPN.
- monitor the traffic and performance of the network, and develop models and provide tools and data for the planning of future networks, especially concentrating on the requirements of grids handling significant volumes of data.
- deal with the distributed security aspects of DataGrid.

PPDG July 2000

L. Price 14

WP 8 HEP Applications

Goal: to exploit the developments of the project to offer transparent access to distributed data and high performance computing facilities to the geographically distributed HEP community

WP 9 Earth Observation Science Applications

Goal: to define and develop EO specific components to integrate the GRID platform and bring GRID-aware application concept in the earth science environment.

Goal: provide a good opportunity to exploit Earth Observation Science (EO) applications that require large computational power and access large data files distributed over geographical archive.

WP 10 Biology Science Applications

Goals:

- Production, analysis and data mining of data produced within projects of sequencing of genomes or in projects with high throughput for the determination of three-dimensional macromolecular structures.
- Production, storage, comparison and retrieval of measures of the genetic expression levels obtained through systems of gene profiling based on micro-arrays, or through techniques that involve the massive production of non-textual data as still images or video.
- Retrieval and in-depth analysis of the biological literature (commercial and public) with the aim of the development of a search engine for relations between biological entities.

PPDG July 2000

L. Price 17

WP 11 Information Dissemination and Exploitation

Goal: to create the critical mass of interest necessary for the deployment, on the target scale, of the results of the project. This allows the development of the skills, experience and software tools necessary to the growth of the world-wide DataGrid.

Goal: promotion of the DataGrid middleware in industry projects and software tools

Goal: coordination of the dissemination activities undertaken by the project partners in the European countries.

Goal: Industry & Research Grid Forum initiated as the main exchange place of information dissemination and potential exploitation of the Data Grid results.

PPDG July 2000

L. Price 18

WP 12 Project Management

Goals:

- Overall management and administration of the project.
- Coordination of technical activity within the project.
- Conflict and resource allocation resolution.
- External relations.

PPDG July 2000

L. Price 19

UK HEP Grid (1)

UK HEP Grid Co-ordination Structure in place

- Planning Group, Project Team, PP Community Committee
- Joint planning with other scientific disciplines

Continuing strong UK Government support

- PP anticipating significant support from current government spending review (results known mid-July)

UK HEP activities centered around DataGrid Project

- Involvement in nearly all workpackages (leadership of 2)
- UK testbed meeting planned for next week
- UK DataGrid workshop planned for mid-July

CLRC (RAL+DL) PP Grid Team formed

PPDG July 2000

L. Price 20

UK HEP Grid (2)

Globus Tutorial / workshop

- led by members of Ian Foster's team
- 21-23 June at RAL
- a key focus in current UK activities

UK HEP Grid Structure

- Prototype Tier-1 site at RAL
- Prototype Tier-2 sites being organised (e.g. JREI funding)
- Detailed networking plans (QoS, bulk transfers, etc.)
- Globus installed at a number of sites with initial tests underway

UK HEP Grid activities fully underway

Global French GRID initiatives Partners

Computing centres:

- IDRIS CNRS High Performance Computing Centre
- IN2P3 Computing Centre
- CINES, centre de calcul intensif de l'enseignement
- CRIHAN centre régional d'informatique à Rouen

Network departments:

- UREC CNRS network department
- GIP Renater

Computing science CNRS & INRIA labs:

- Université Joseph Fourier
- ID-IMAG
- LAAS
- RESAM
- LIP and PSMN (Ecole Normale Supérieure de Lyon)

Industry:

- Société Communication et Systèmes (CS-SI)
- EDF R&D department

- Applications development teams (HEP, Bioinformatics, Earth Observa
- IN2P3, CEA, Observatoire de Grenoble, Laboratoire de Laplace



Data GRID Resource target

	2001	2002	2003
CPU (SI95)	4,000	14,000	60,000
Disk (TB)	4	12	50
Robotic (PB)	0.2	0.3	1
Lyon-CERN Mbs	34	622	2,500
Tier1-Tier2 Mbs	34	155	1,000

•France (CNRS-CEA-CSSI)
•Spain (IFAE-Univ. Cantabria)

PPDG July 2000 L. Price 23

Dutch Grid Initiative

- NIKHEF alone not "strong" enough
- SARA has more HPC infrastructure
- KNMI is the Earth Observation partner
- Other possible partners
- Surfnet for the networking
- NCF for HPC resources
- ICES-KIS for human resources
- ..

PPDG July 2000 L. Price 24

Contributions to WPs

Funded

Work package		NIKHEF (CR5)	SARA (AC19)	KNMI (AC18)
WP4	Fabric Man.	24		
WP6	Test bed		6	
WP7	Network Man.		18	
WP8	HEP Applications	12		
WP9	Earth Observation			24
Total		36	24	24

Un-funded

Work package		NIKHEF (CR5)	SARA (AC19)	KNMI (AC18)
WP2	Data Man.		36	
WP4	Fabric Man.	72		
WP5	Mass Storage Man.		36	
WP8	HEP Applications	54		
WP9	Earth Observation			24
Total		126	72	24

PPDG July 2000

L. Price 25

Initial Dutch Grid Coll.

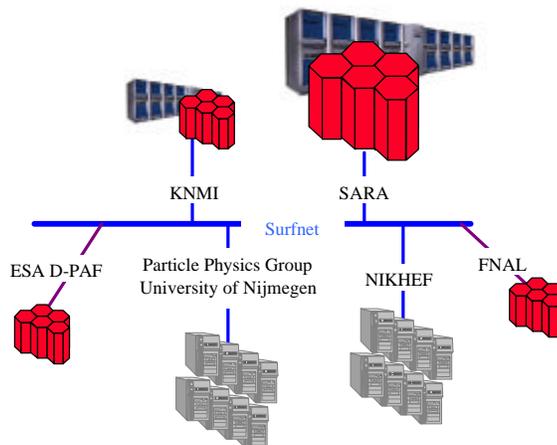
**NIKHEF and SARA and KNMI and
Surfnet and NCF and GigaPort and ICE-KIS and ...**

**Work on Fabric Man. Data Man. and Mass Storage
And also (later?) on Test bed and HEP applications**

PPDG July 2000

L. Price 26

Initial Dutch Grid topology



PPDG July 2000

L. Price 27

Grid project 0

On a few (distributed) PC's

- Install and try GLOBUS software
- See what can be used
- See what needs to be added

Study and training

Timescale: from now on

PPDG July 2000

L. Price 28

Grid project I

For D0 Monte Carlo Data Challenge

- Use the NIKHEF D0 farm (100 cpu's)
- Use the Nijmegen D0 farm (20 cpu's)
- Use the SARA tape robot (3 Tbyte)
- Use the Fermilab SAM (meta-) data base

Produce 10 M events

Timescale: this year

Grid project II

For GOME data retrieving and processing

- Use the KNMI SGI Origin
- Use the SARA SGI Origin
- Use the D-PAF*) data center data store
- Use experts at SRON and KNMI

(Re-) process 300 Gbyte of ozone data

Timescale: one year from now

*) German Processing and Archiving Facility

And thus portal to CERN and FNAL and ...

High bandwidth backbone in Holland

High bandwidth connections to other networks

Network research and initiatives like GigaPort

(Well) funded by the Dutch government

INFN-GRID and the EU DATAGRID

Collaboration with EU DATAGRID on:

- common middleware development
- testbed implementation for LHC for Grid tools tests

INFN-GRID will concentrate on:

- Any extra middleware development for INFN
- Implementation of INFN testbeds integrated with DATAGRID
- Computing for LHC experiments for the next 3 years
 - Prototyping Regional Centers Tier1 ...Tiern
- Develop similar computing approach for non-LHC experiments such as: Virgo and APE
- Incremental development and test of tools to satisfy the computing requirements of LHC and Virgo experiments

PPDG July 2000

L. Price 31

Direct participation in EU DATAGRID

Focus on Middleware development (first 4 WPs) and testbeds (WP6-7) for validation in HEP ..and in other sciences WP9-11
L'INFN responsible for WP1 and participate in WP2-4,6,8,11

PPDG July 2000

L. Price 32

Eurogrid

HPC centres

- IDRIS (F)
- FZ Juelich (D)
- Parallab (N)
- ICM (PL)
- CSAR (UK)

Users

- German Weather service (D)
- Aerospatiale (F)
- *debis Systemhaus (D) (assistant partner)*

Technology integrators

- Pallas (D)
- *Fecit (UK) (assistant partner)*

PPDG July 2000

L. Price 33

Strategic position of Eurogrid

GRID of leading HPC centres

Application-oriented

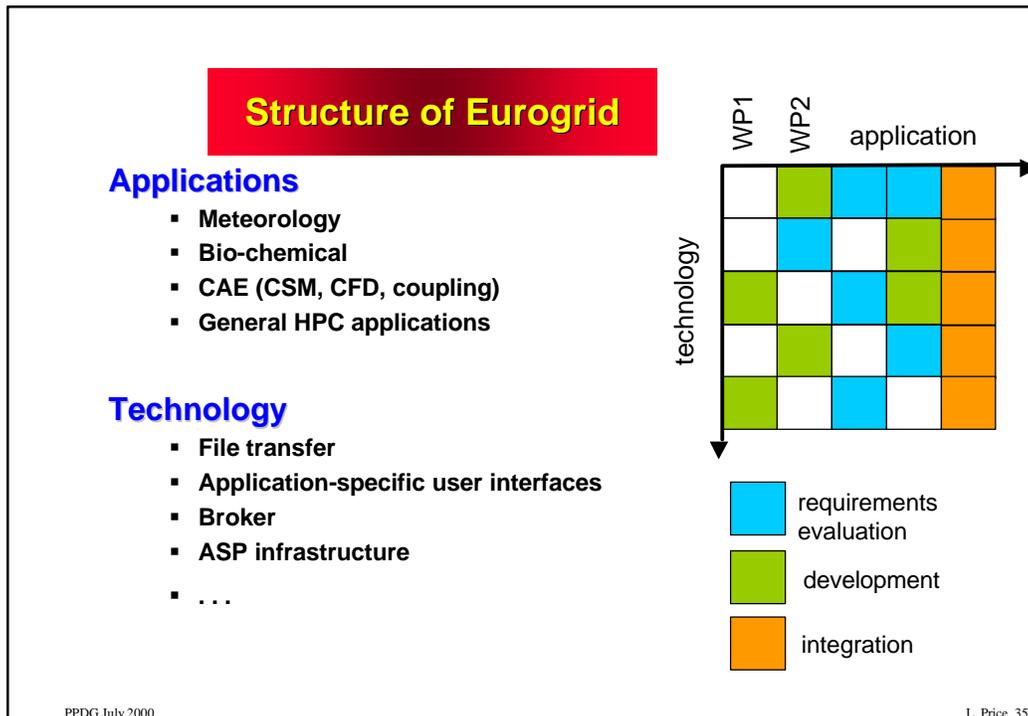
Industrial participation (e-business components)

**Starting with existing advanced technology
(UNICORE)**

**Open for cooperation with other projects
(Globus, ..., EU Datagrid)**

Exploitation as product planned

PPDG July 2000



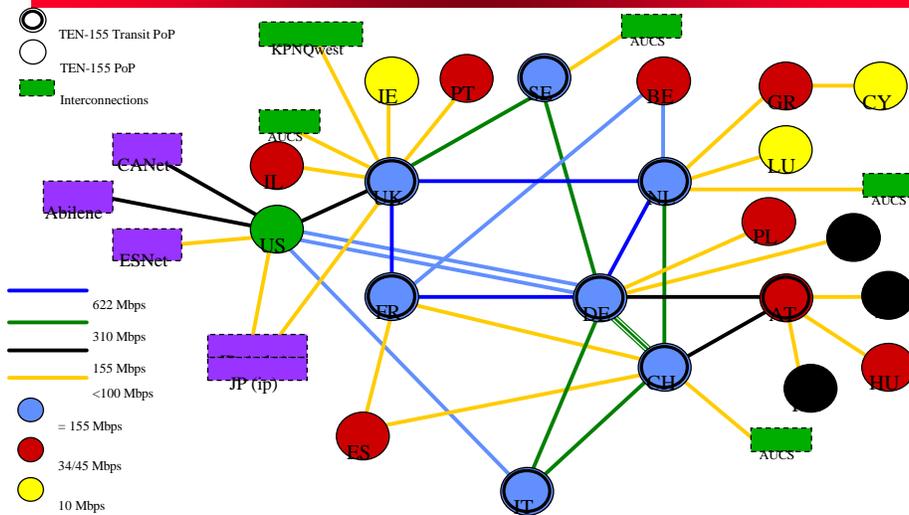
EU DataGrid Main Issues

- Project is by EU standards very large in funding and participants**
- Management and coordination will be a challenge**
- Coordination between national and DataGrid programmes**
- Coordination with US Grid activity**
- Coordination of HEP and other sciences objectives**
- Very high expectations already raised (too soon?) could bring to (too early?) disappointments**
- EU Project not approved (yet?)**

PPDG July 2000

L. Price 37

TEN-155 Topology (2000)



PPDG July 2000

L. Price 38

European Infrastructure (1)

Multi-Gigabit Core

- 2÷10 Gb/s in 2001
- Single or Multiple 1's
- Lit Fibers



Core Will Expand in Time

- 6-10 Locations in 2001
- 20 Locations by 2003

Capacity Will Increase Yearly 2÷4 Times

PPDG July 2000

L. Price 39

European Infrastructure (2)

Non-Core Locations

- 34-622 Mb/s in 2001
- Multiply Connected
- SDH (+ ATM ?)



Capacity Will Increase Yearly 2÷4 Times

Will Migrate to Core as Soon as Technically and Commercially Possible

PPDG July 2000

L. Price 40

European Distributed Access (EDA)



Rationalize Research Connections With Other World Regions

- N.–America, Asia–Pacific, S.–America, ...

Connections May Be Activated at Any PoP of the Core

Transparent Access To/From Any Other PoP

PPDG July 2000

L. Price 41

Research Support

Extensive VPN capability

- ATM
- MPLS
- I Allocation



Guaranteed Capacity Service (GCS)

Premium–IP service

Best Effort IP

PPDG July 2000

L. Price 42

Grids

Capacity Requirements

- In Line with GEANT Specifications

VPN Requirements

- Can Be Fulfilled if NREN's Can Support It

Funding Support

- Lobbying at the NREN Level Could Be Needed



Budget Estimate

~ 210 MEuro in 4 Years

- 62% from NREN's
- 38% from EC

EDA Infrastructure is Included

Intercontinental Budget and Support is Being Negotiated



GEANT Timetable

- EC Approves Project GN1 June 21st, 2000
- EC Contract Negotiations July 2000
- Tender Expiration September–End, 2000
- Evaluation October–November 2000
- Funding for TEN-155 Expires October–End, 2000
- GN1 Starts November 1st, 2000
- Delivery of GEANT Infrastructure Summer 2001
- Overlap End-Time October 31st, 2001



PPDG July 2000

L. Price 45

How do we Integrate these Projects?

Multiple interlocking projects and goals

- Merger of CS research and mission requirements

Physics (Science?) Grid Steering (Coordination?) Committee

- Attempt to partition research goals
- Agree on interoperability standards
- Coordinate testbed plans
 - Partition emphasis where possible
 - Interoperation is vital
 - Plan capability vs time

Role of Grid Forum?

PPDG July 2000

L. Price 46