

Guy Wormser
IN2P3/CNRS ,
EGEE Applications Manager

December 18, 2003



EGEE

Enabling Grids for E-science in Europe



EGEE is proposed as a project funded by the
European Union under contract IST-2003-508833

EGEE : Enabling Grids for E-science in Europe

Enabling Grids for
E-science in Europe

Goal

create a general European Grid
production quality infrastructure on top of
present and future EU RN infrastructure

Build on

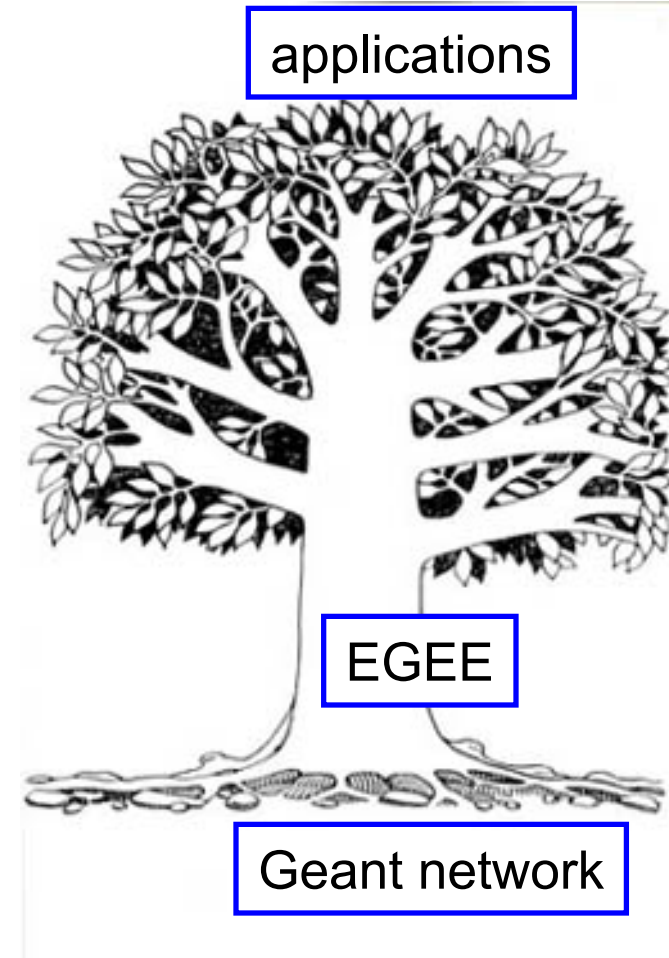
EU and EU member states major
investment in Grid Technology
Several pioneering prototype results
Largest Grid development team in the
world

Goal can be achieved for about €100m/4 years on top
of the national and regional initiatives

Approach

Leverage current and planned national
and regional Grid programmes (e.g.
LCG)

Work closely with relevant industrial Grid
developers, NRNs and US



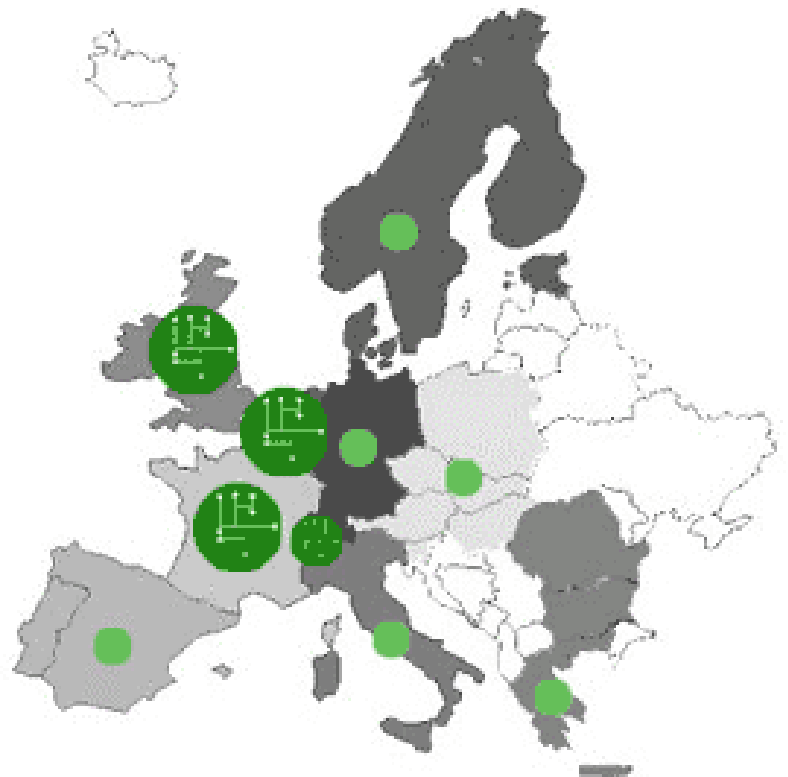
EGEE vision

- **Boost European research and industry** by providing easiest access to very large computing resources using GRID technologies
 - Satisfy the needs of well identified demanding scientific applications
 - Open up new horizons for a variety of other academic and industrial users
- Build on top of existing tools (« hit the ground running »)

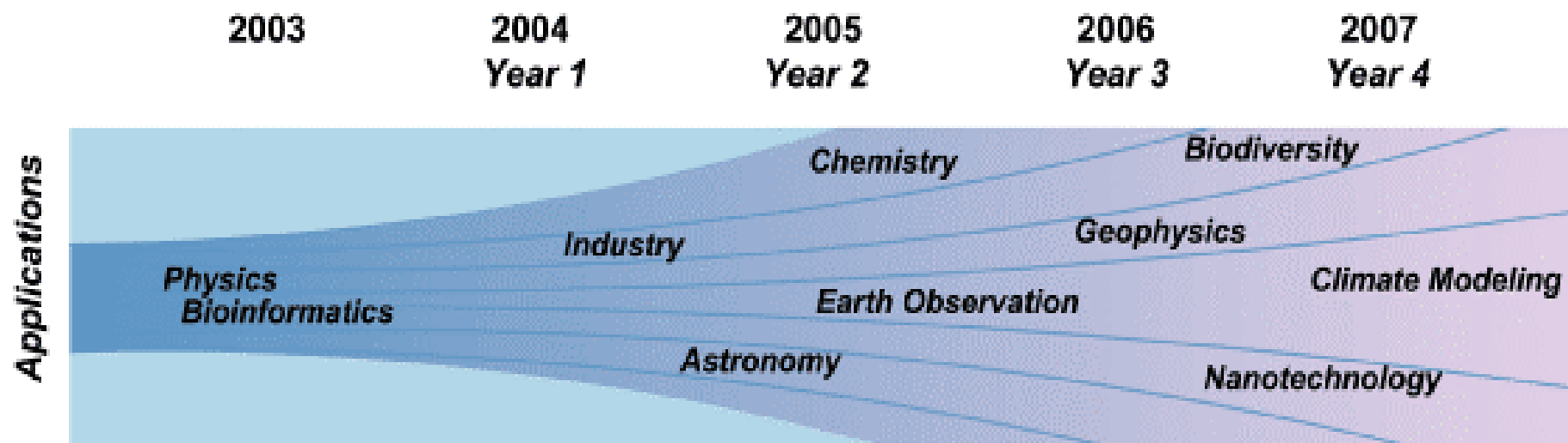
EGEE Applications Sector

- EGEE applications scope
 - The Pilot applications
 - The Generic Applications
- The applications coordination
- The testing team
- NA4 structure
- Conclusion

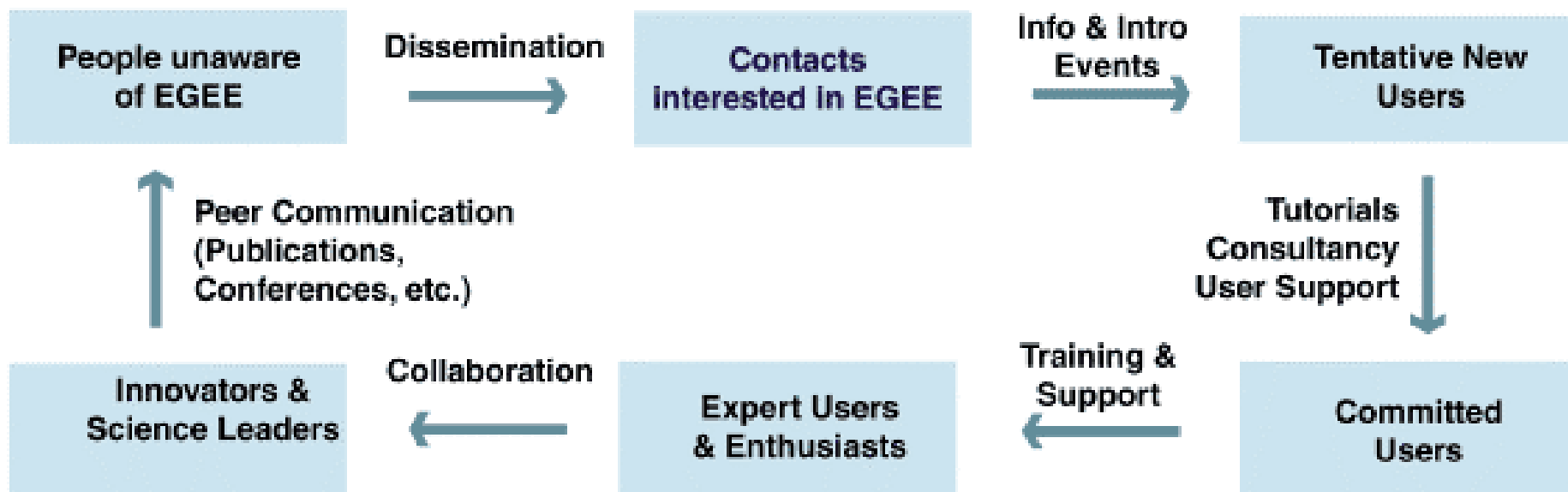
Applications, Dissemination, User Training



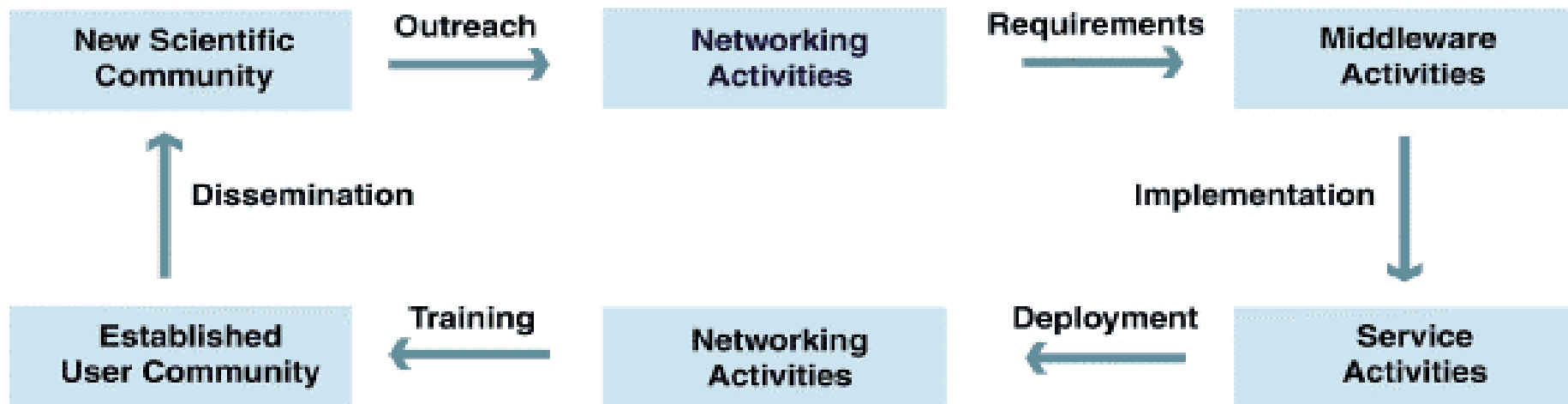
-  Lead Networking Centre
-  Pilot Application Centre
-  Regional Networking Centre



EGEE virtuous circle (1)



EGEE « virtuous circle » (2)



EGEE Applications scope

- EGEE Scope : ALL-Inclusive for academic applications
- Open to industrial and socio-economic world as well
- The **only success criterium** of EGEE!: how many satisfied users from how many different domains ?
- In the TA: **5000 users** (3000 after year 2) from **at least 5 disciplines**
- Firmly establish the added value brought by the Grid on a
 - **quantitative** basis (« much more of the same »)
 - **qualitative** basis (« breaking new grounds thanks to the Grid concept »)

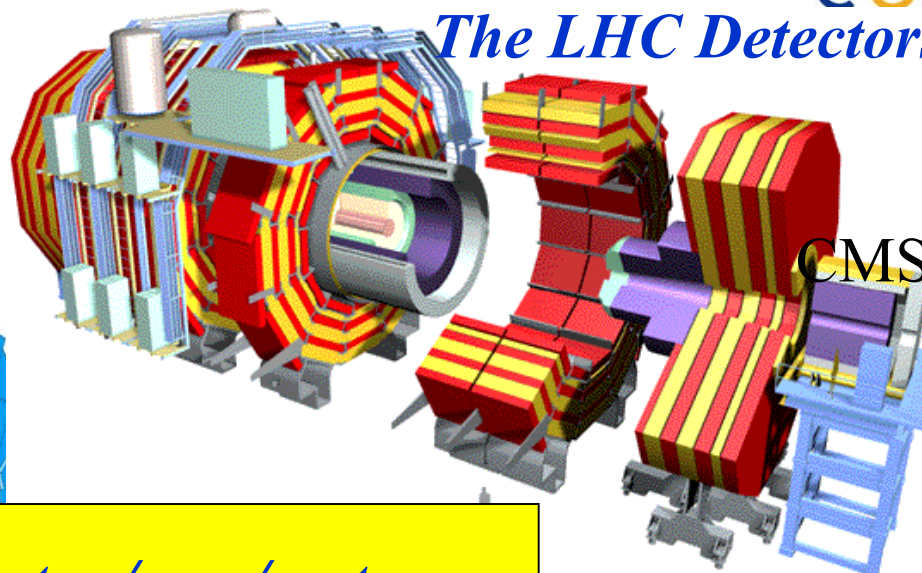
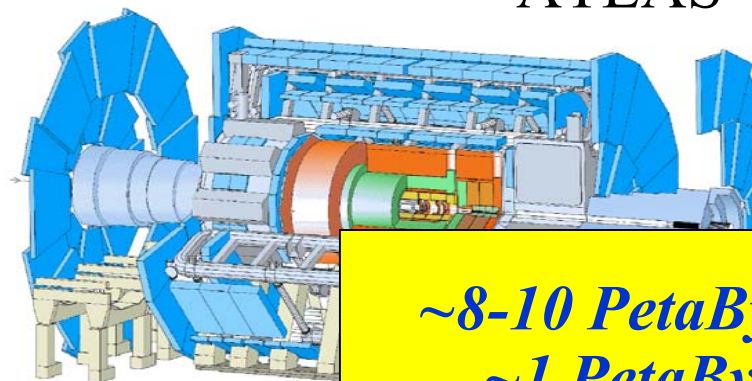
The pilot applications

- Dual role
 - Earliest users
 - Feedback providers
- Must be dedicated and grid-aware **communities**
 - **HEP and Bio/medical**
- Natural continuation from DATAGRID

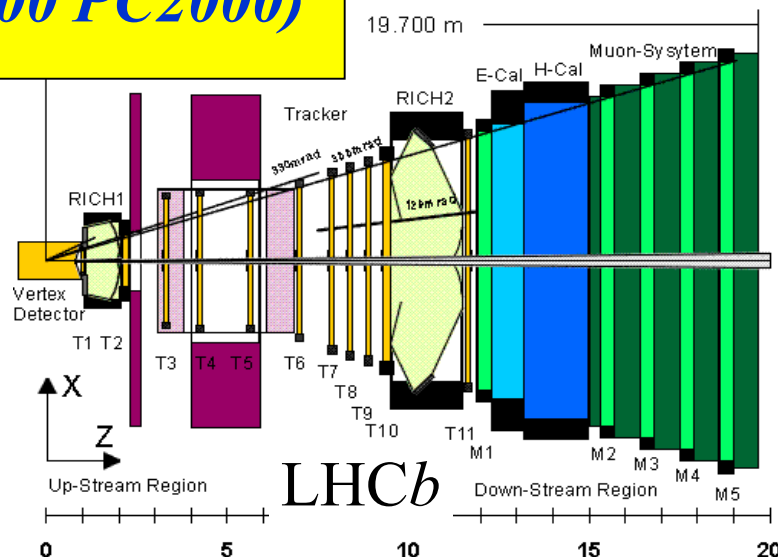
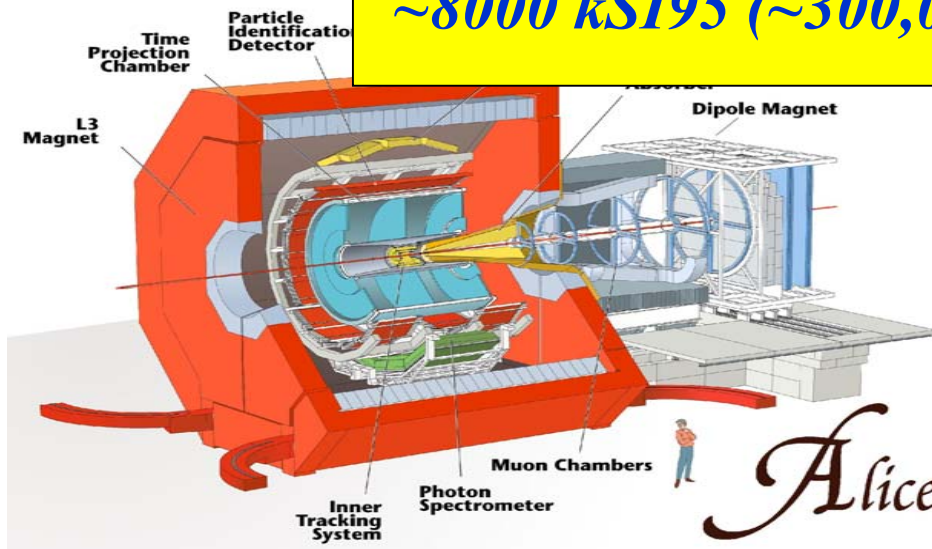
The LHC Detectors

ATLAS

CMS



*~8-10 PetaBytes /year/on tape
~1 PetaByte/year/on disk
~8000 kSI95 (~300,000 PC2000)*



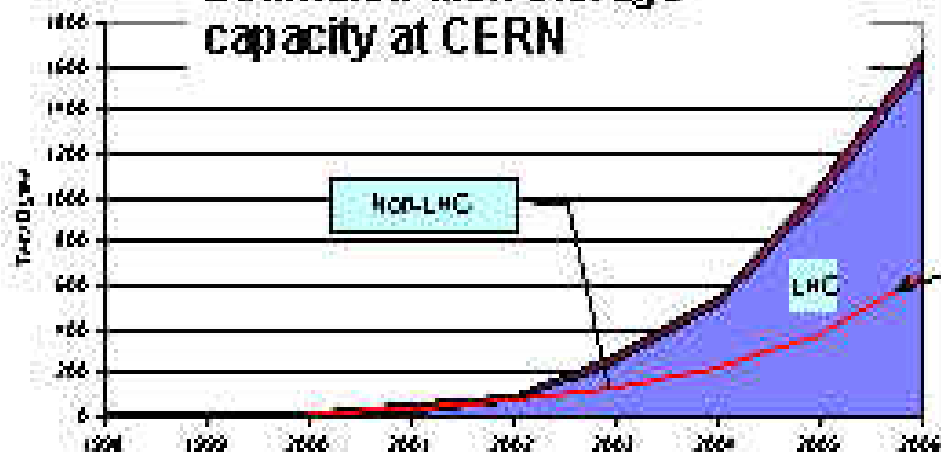
CERN's Network in the World



Europe: 267 institutes, 4603 users
Elsewhere: 208 institutes, 1632 users

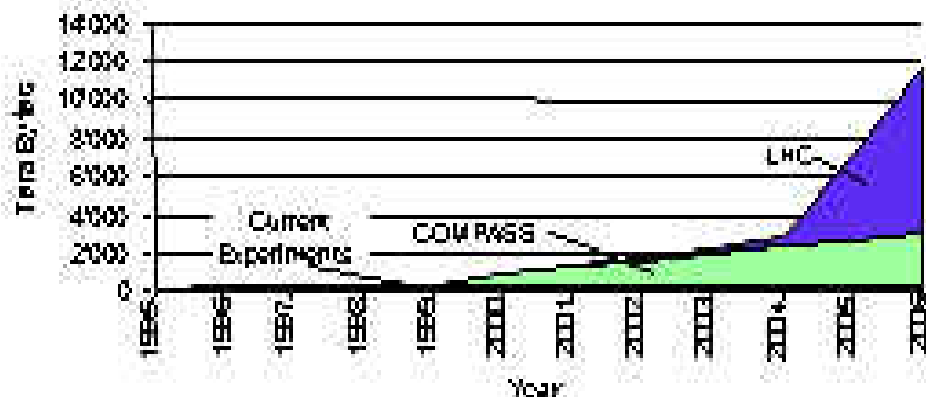
New solutions are necessary!

Estimated disk storage capacity at CERN

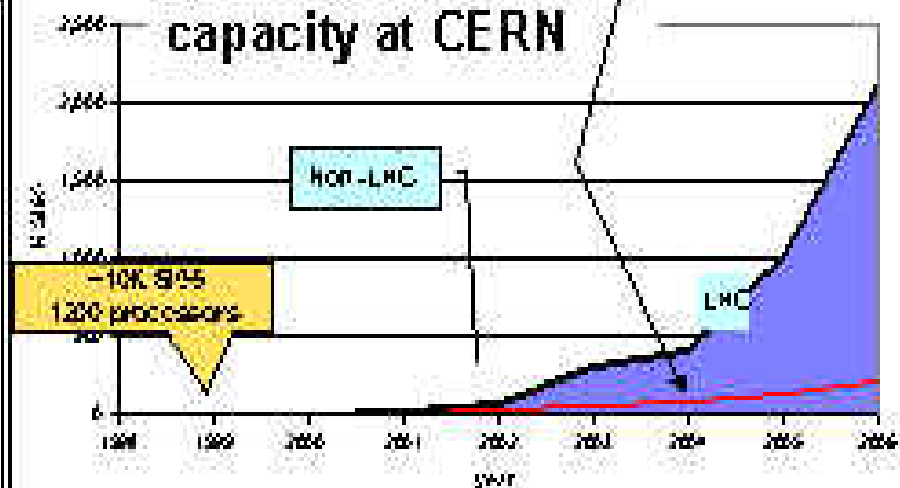


Small fraction of the main analysis capacity will be at CERN

Long Term Tape Storage Estimates



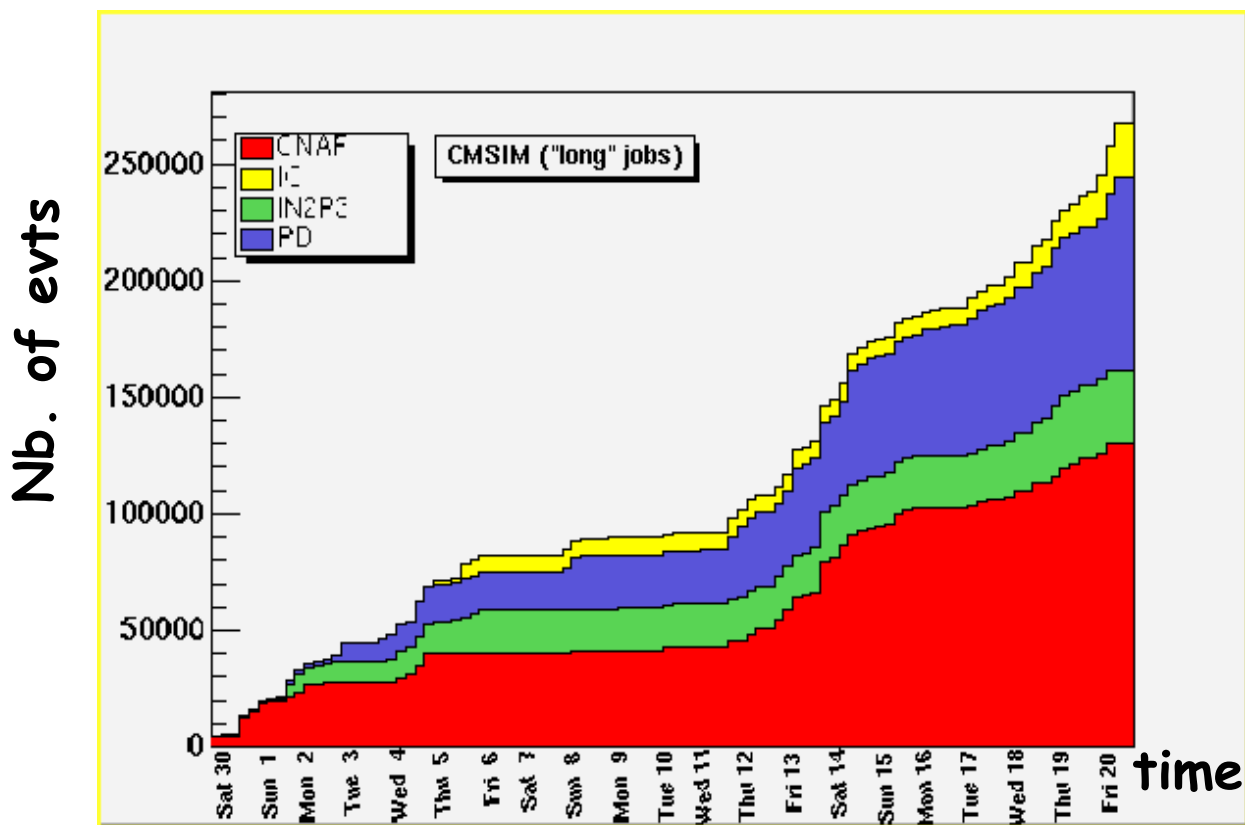
Estimated CPU capacity at CERN



Moore's Law

CMS event production in December 2002 using EDG software and applications TB

<http://cmsdoc.cern.ch/cms/production/www/html/general/index.html>



Application Testbed Resources

Since Last Year:

- Improved software (EDG 1.4.3)
- Doubled sites. More waiting...
 - Australia, Taiwan, USA (U. Wisc.), UK Sites, INFN, French sites, CrossGrid, ...
- Significantly more CPU/Storage.

Hidden Infrastructure

- MDS Hierarchy
- Resource Brokers
- User Interfaces
- VO Replica Catalogs
- VO Membership Servers
- Certification Authorities

Site	Country	CPU	Storage
CC-IN2P3*	FR	620	192 GB
CERN*	CH	138	1321 GB
CNAF*	IT	48	1300 GB
Ecole Poly.	FR	6	220 GB
Imperial C.	UK	92	450 GB
Liverpool	UK	2	10 GB
Manchester	UK	9	15 GB
NIKHEF*	NL	142	433 GB
Oxford	UK	1	30 GB
Padova	IT	11	666 GB
RAL*	UK	6	332 GB
SARA	NL	0	10000+
TOTAL	5	1075	14969 GB

*also Dev. TB; +200 TB including tape

Application Testbed Users

VO	User
CMS	106
WP6	87
ALICE	63
ATLAS	55
Earth Obs.	29
BaBar	29
LHCb	28
ITeam	22
Genomic	22
TSTG	16
Medical	6
Doc.	3

Certificate Authorities Group

- Evaluates & approves new CAs
- 16 currently approved.
- Collaborating w/ other grid proj.
- More on the way...
 - Cyprus
 - US FNAL (KCA)
 - Belgium
 - Taiwan



CA	User
INFN (IT)	113
CNRS (FR)	71
UK	58
CERN (CH)	44
NIKHEF (NL)	19
Russia	15
US DOE	10
Spain	8
FZK (D)	5
Czech Rep.	3
Portugal	3
NorduGrid	2
Poland	1
Canada	0
Greece	0
Slovakia	0
TOTAL	352

Virtual Organizations

- Also for Storage Elements
- Guidelines (EDG rules)

Course-grained Authorization.



CMS/EDG Summary of Stress Test

Preliminary Analysis

Short jobs

After Stress
Test – Jan 03

CMKIN jobs			
Status	EDG evaluation	CMS evaluation	EDG ver 1.4.3
Finished Correctly	5518	4601	604
Crashed or bad status	818	1099	65
Total number of jobs	6336	5700	669
Efficiency	0.87	0.81	0.90

Long jobs

After Stress
Test – Jan 03

CMSIM jobs			
Status	EDG evaluation	CMS evaluation	EDG ver 1.4.3
Finished Correctly	1678	2147	394
Crashed or bad status	2662	934	104
Total number of jobs	4340	3081	498
Efficiency	0.39	0.70	0.79

HEP sector

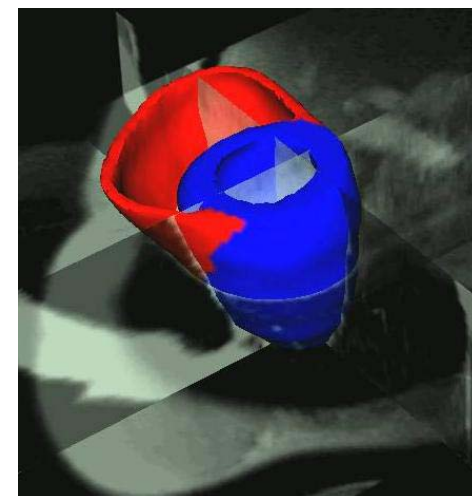
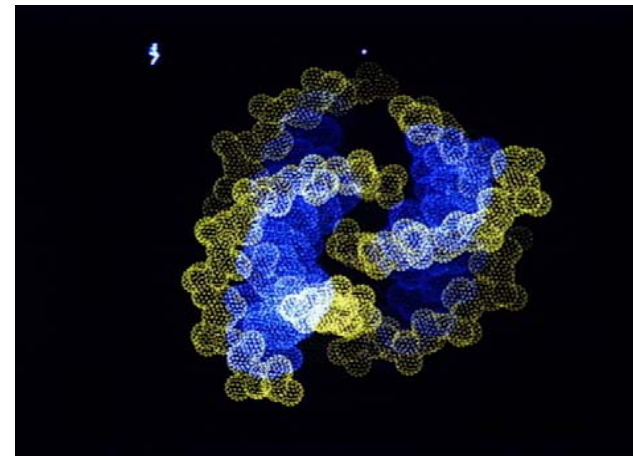
- CERN coordinator
 - 4 FTE plus 4 unfunded
 - Embedded within LHC experiments
- Role : deployment of LHC apps on EGEE infrastructure and feedback
- Open to non-LHC experiments as well

Bio sector

- CNRS coordinator, Spain and Russia partners
 - 4 FTE funded + 4 unfunded
 - Embedded in the BioMed projects, on a permanent or temporary basis
- Multiple role:
 - Demonstrate the usefulness of grid-powered biomed applications
 - Become the focal point of the Biomed Grid community, in partnership with the HealthGrid association
 - Provide detailed feedback to EGEE
 - Ensure that specific bioMed requirements propagate everywhere in EGEE

Biomedical applications

- Data mining on genomic databases (exponential growth).
- Indexing of medical databases (Tb/hospital/year).
- Collaborative framework for large scale experiments (e.g. epidemiological studies).
- Parallel processing for
 - Databases analysis
 - Complex 3D modelling

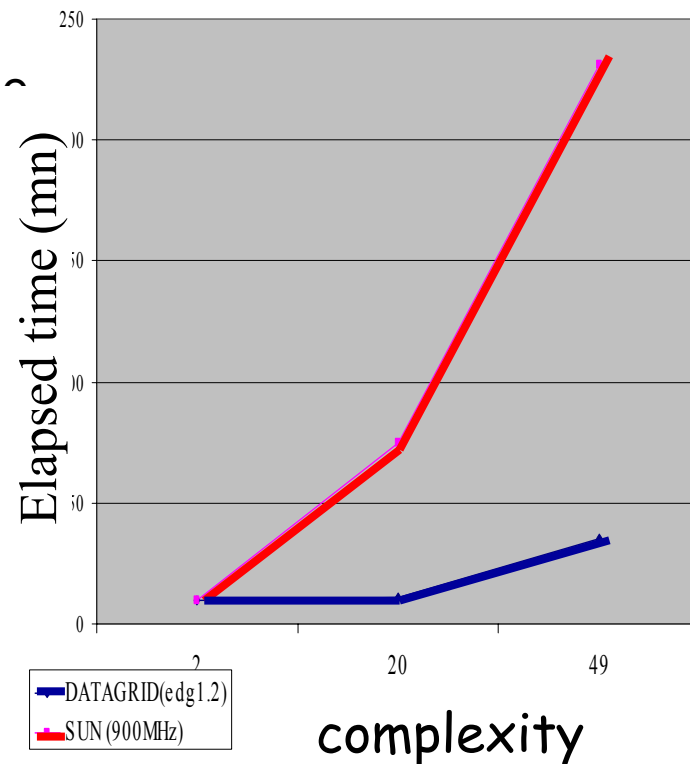


Addressing the key challenges...

- More successes...

- More than 15 WP10 users trained
- Several system administrators initiated to the installation procedure of a grid node
- First evaluation of performance

Phylogenetics (fastDNAm algorithm)



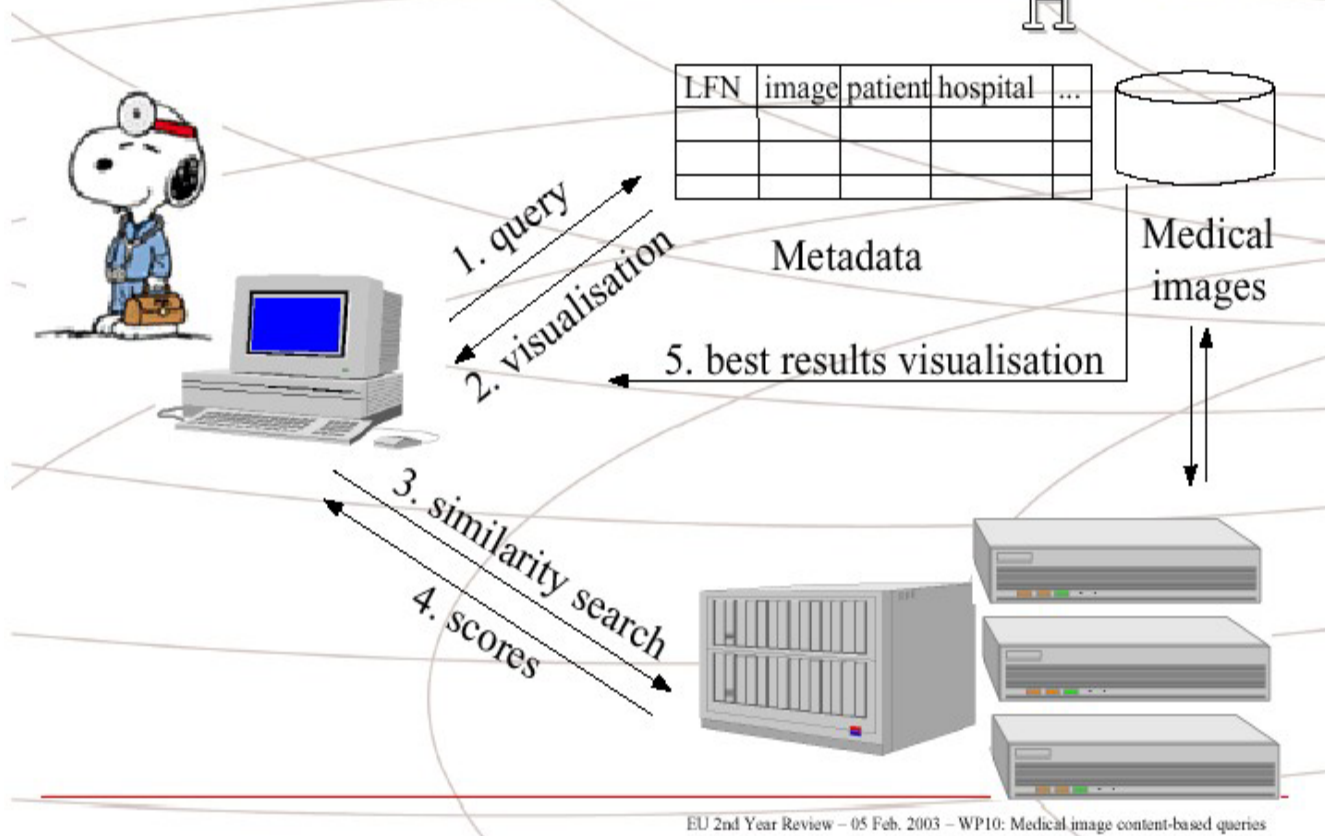
Credit : T. Silvestre BBE



Application synopsis



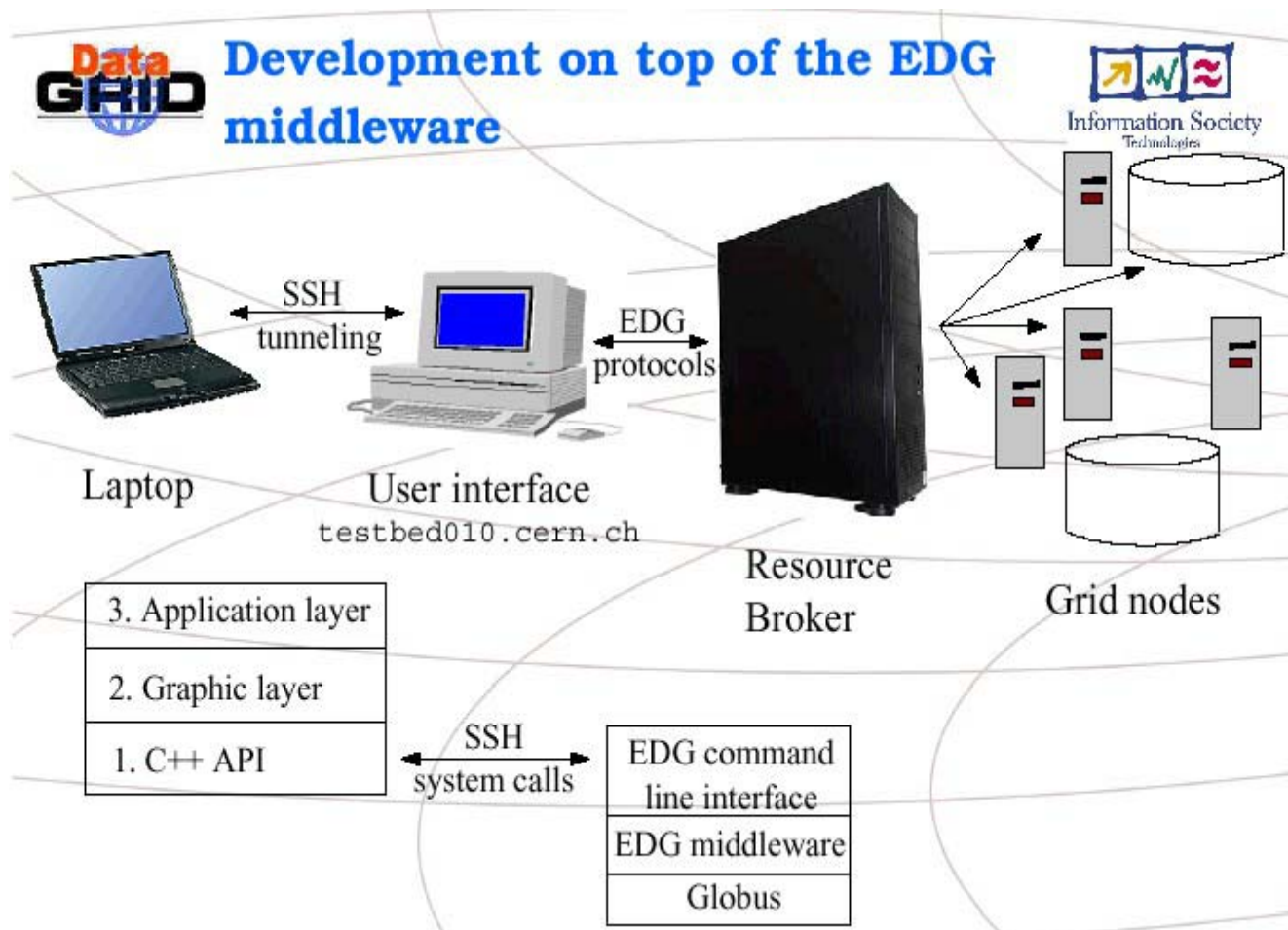
Information Society
Technologies



EU 2nd Year Review – 05 Feb. 2003 – WP10: Medical image content-based queries



Development on top of the EDG middleware



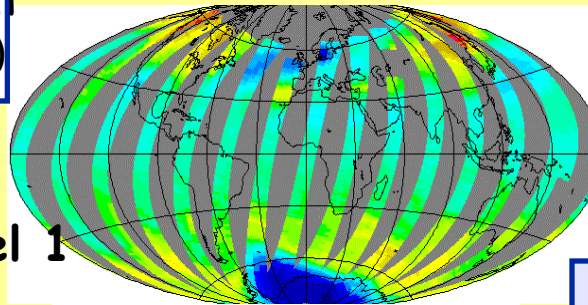
EO Use Case File Numbers

1 Year of GOME data :

Dataset	Number of files	Size
Level 1	4,724	15 Mb
Level 2 NNO (ESA) Level 2	9,448,000	10 kb
Opera Kermit Kermit	9,448,000	12 kb
Lidar (IPSL)	12	2.5 Mb
Total:	18,900,736 files	267 Gb

(just one part of a 5-year dataset)

Raw satellite data from the GOME instrument (~75 GB - ~5000 orbits/y)



(example of 1 day total O₃)

Level 1

ESA(IT) KNMI(NL)
Processing of raw GOME data to ozone profiles.
2 alternative algorithms
~28000 profiles/day



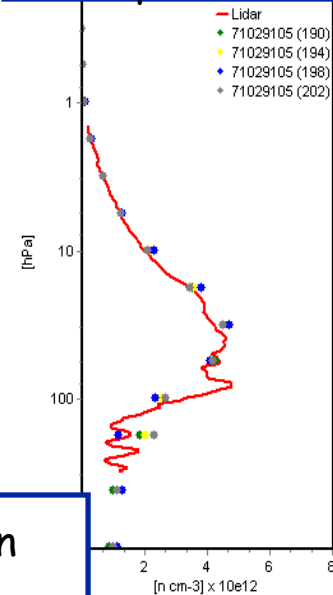
Level 2

IPSL(FR)
Validate some of the GOME ozone profiles (~10⁶/y)
Coincident in space and time with Ground-Based measurements

DataGrid environment

Visualization & Analyze

LIDAR data (7 stations, 2.5MB per month)



The generic applications

- All the others!
- Need of an attractive and orderly integration process:
 - Peer review process
 - scientific interest of the proposed work, with particular emphasis on the grid added-value,
 - coordination of the corresponding community,
 - grid-awareness of this community
 - minimum requirement that a small team followed the EGEE training), dedication of the community to this application,
 - agreement to the various EGEE policies and especially the security and resources allocation policies.
- Technical involvement
 - Allocate workforce for a given time period
- Ressource allocation policy
 - Initial period: free lunch, then roughly-diagonal model

Applications coordination

- Each application sector has operational autonomy
 - coordinator
 - dedicated manpower
 - specific goals
 - Seat in EGEE Architecture group
- However, clear need for overall coordination (cf WP8-9-10 working group)
 - Common tools
 - Cross fertilization
 - Unified user interface
 - relationship with other EGEE sectors (PEB seat)

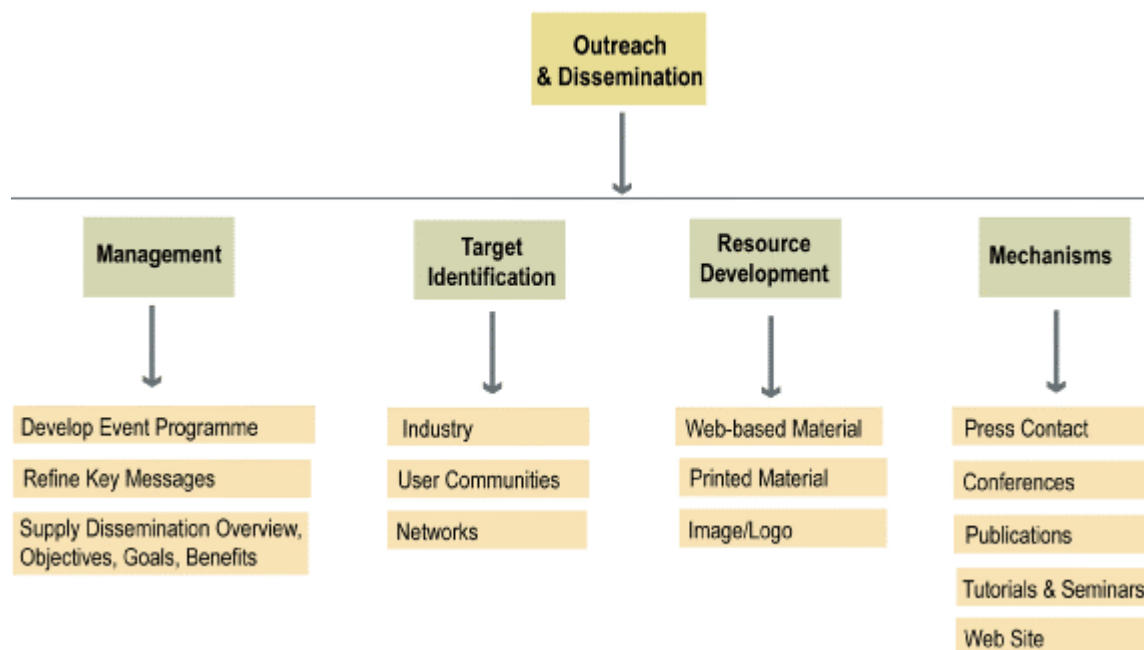
Other NA4 activities

- Industry Forum
 - Presession April 22, 2003 (~25 participants)
 - Second session October 7 , 2003 at CERN
- NA3 liason person
- NA4 Test team
 - Derive testing suites based on use cases

NA4 Steering committee

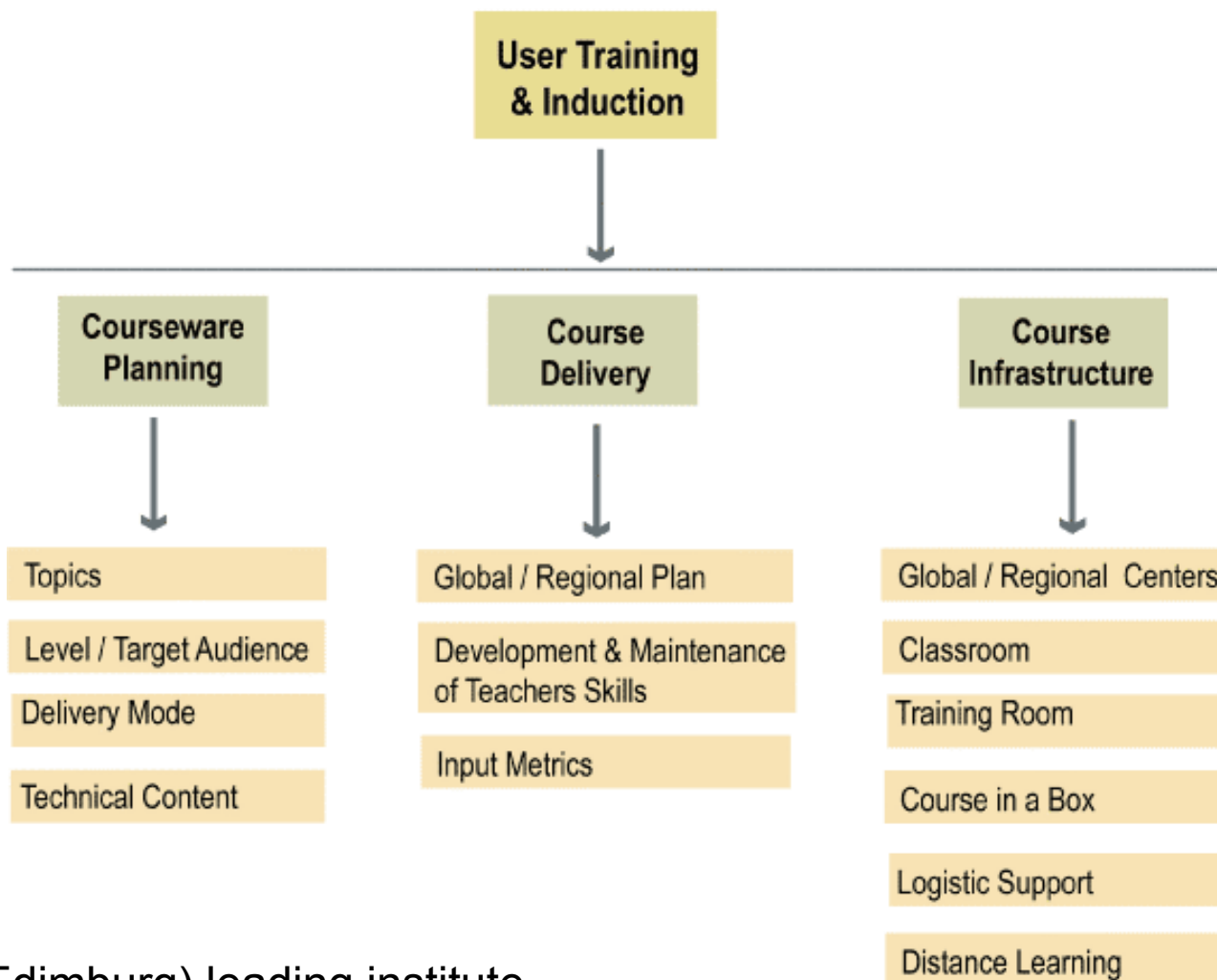
- Guy Wormser, CNRS-France, NA4 overall manager.
- Vincent Breton , CNRS-France, NA4 Bio-Med Application manager
- Frank Harris, CERN, NA4 HEP Application manager.
- Roberto Barbera, INFN-Italy, NA4 Generic Application manager.
- Christian Saguez, Ecole Centrale Paris-France, Chair of the EGEE Industry Forum.
- Francois Etienne, CNRS-France, NA4 deputy manager.

Dissemination and outreach (NA2)



TERENA leading institute

User training and Induction (NA3)



NeSC (U. Edimburg) leading institute

Summary: How to join EGEE and beyond

- Get in touch through dissemination activities
 - (Already done for you!
- Contact NA4 group and start the process
- Think about what will be the grid-added value for you
- What will the required help needed from EGEE in terms of manpower and what are your own resources?
 - Train your interface team using the EGEE training courses
- What will be the required computing resources from EGEE and what are your own resources
 - Think about the evolution between the two
- Fill the questionnaire and pass the EGEE selection process
- Start deploying your application!

Resource Allocation Policy

- Manpower
- Allocate a team out of the EGEE pool of 8 people for a limited period (a few months) to help you
 - Interface your applications with EGEE software
 - Deploy on EGEE infrastructure
 - Run it
- After some initial period, contact points thru EGEE User Support mechanism
- Computing resources
- Initial free use of the EGEE infrastructure
- Bring your own resources and incorporate them into the EGEE infrastructure
- Benefit from EGEE resources and give a fraction of yours

Conclusion

- The Application sector is at the end of EGEE chain
- But it is the **ONLY success criterion**:
 - **Many happy users from many different fields (academic and some industry-related)**
- Two pilot applications with special role to provide detailed feedback: **HEP and BioMed**
- **Well defined integration process** to « adopt » new applications
- Small workforce dedicated to one application at a time to help its deployment
- General application coordination to progress towards a unified interface layer
- Interface with all EGEE areas