

## The word '*grid*' is (over)used a lot (HYPE)

- Oracle databases ;(
- cluster computing
- cycle scavenging
- “If a customer calls it a ‘grid’, then it is a grid”
- **cross-domain resource and data sharing**

## Is there a clear definition?

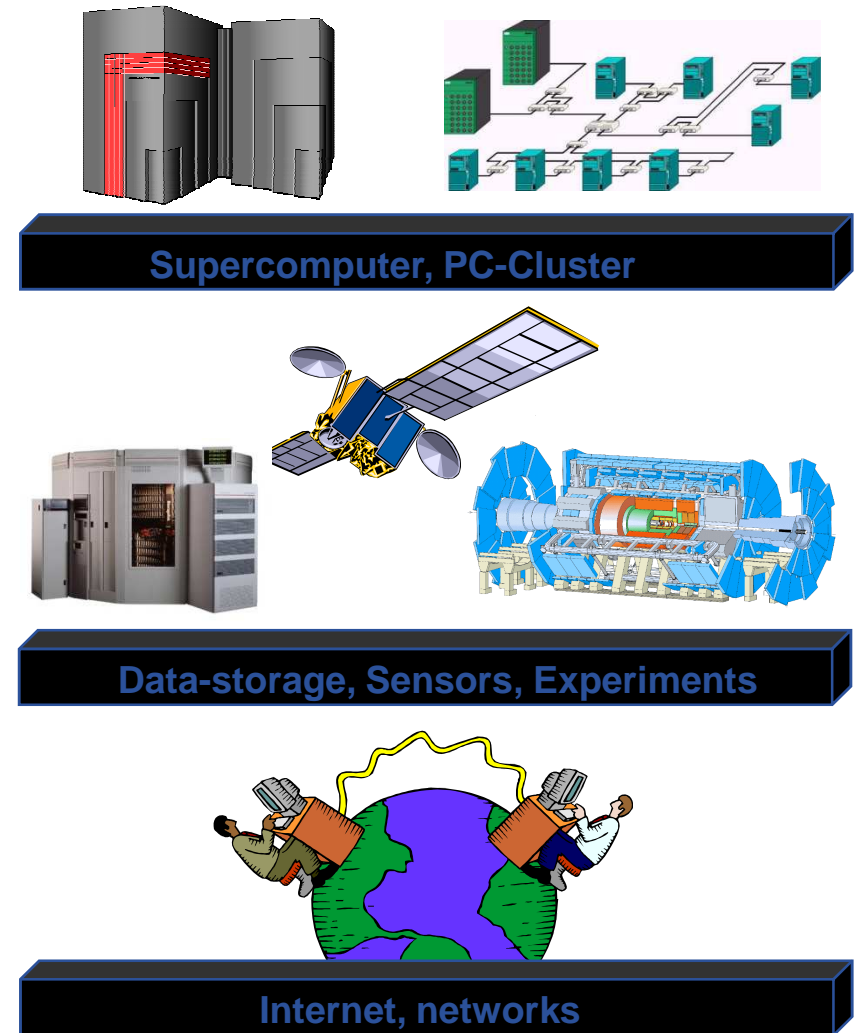
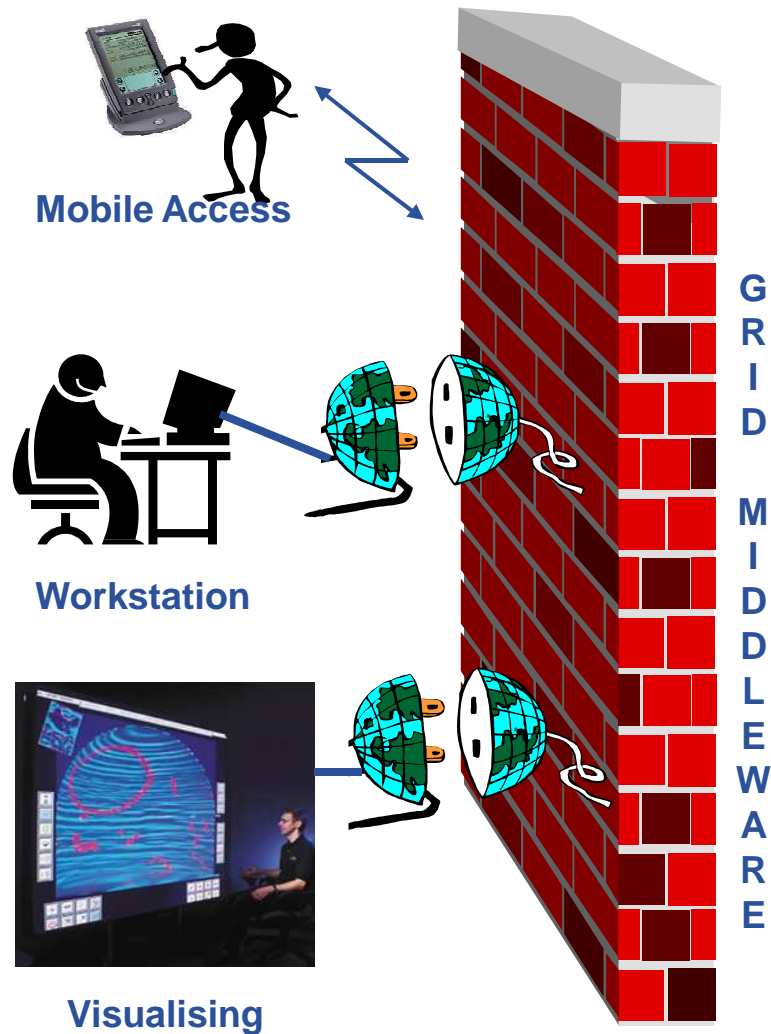
- Coördinate resources not under a central controle
- The use of standards, open and generic protocols & interfaces
- Delivering a non-trivial amount of collective services

### When do you need a grid?

- More then one computer
- More then one use (sharing)
- More then one location (collaborating)
- More then one company/devision
- More then one community

In general: More then ONE.

# The Grid metaphor





Enabling Grids for E-scienceE

# Introduction GRID Tutorial

Maurice Bouwhuis  
SARA

[www.eu-egee.org](http://www.eu-egee.org)

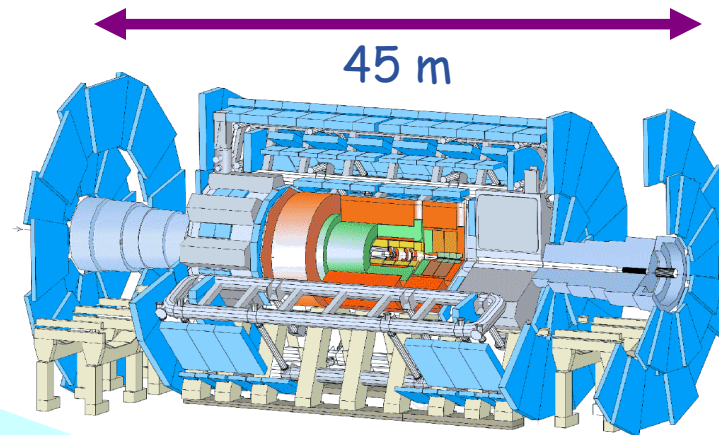


# Problem 1: Hoge Energie Fysica

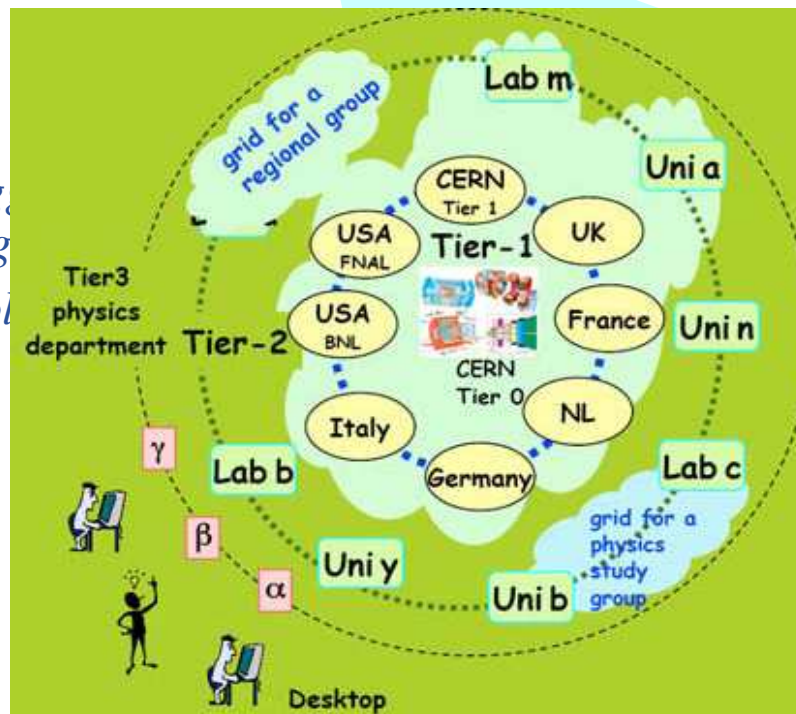


## ATLAS

*One of the four  
LHC detectors*



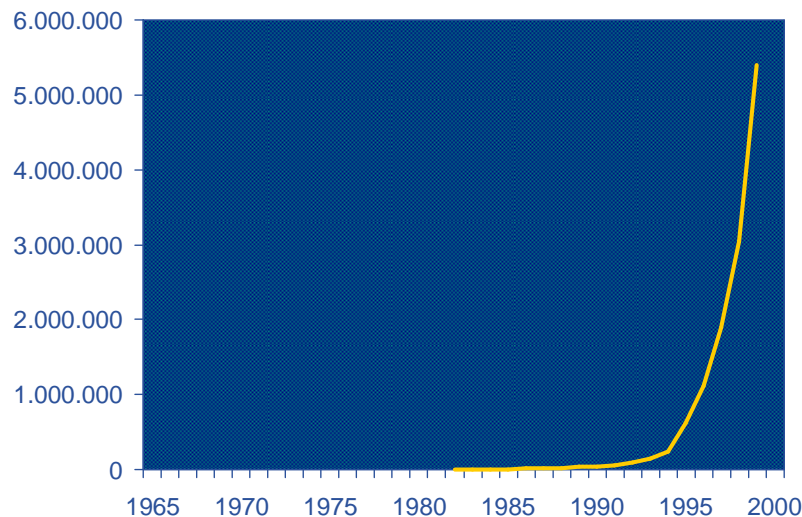
**online system**  
*multi-level trigger  
filter out background  
reduce data volume*



In NL:

- ~3000 CPU
- 3.000.000 GB disk
- 3.000.000 GB tape per jaar

- Autometisering en increased resolution → more data AND more complex data
- Many different sources
- Hypothesis driven → data driven



Over 5 million  
sequence entries in  
GenBank

Over 3 billion  
bases from  
41,000 species

- **Astronomy (LOFAR et al)**
- **Climate research**
- **Earth observation**
- **Alpha en Gamma sciences**
  - Storage and Long Term archiving
  - Analysis of digital files
- **Ecology**
- **Food and health**
- **Medical instrumentation design**
- **Archaeology**
- .....

Where are you from

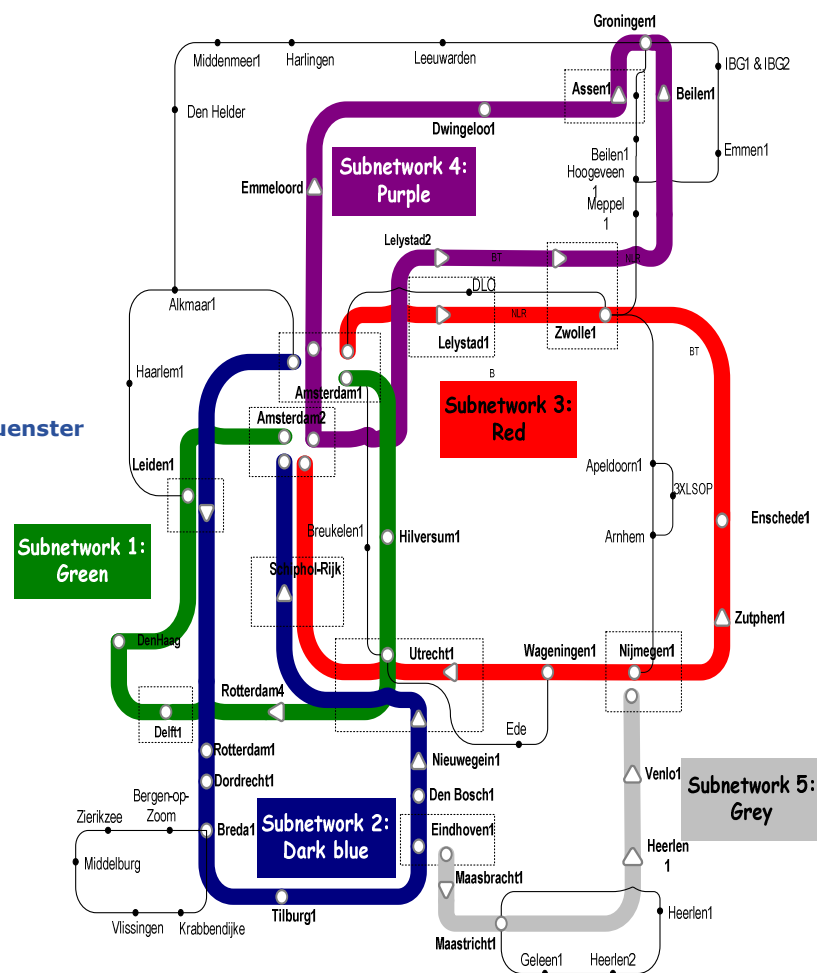
- **So what is happening today?**
  - Scale! Grid infrastructures operate worldwide
    - International infrastructures - EGEE, DEISA, Nordugrid, OSG, TeraGrid
    - National – NAREGI (Japan), UK-eScience, D-Grid, NLGrid
  - Interoperability – availability of middleware – Globus toolkit, UNICORE, NAREGI, schedulers

- **Some basic requirements for a grid infrastructure**
  - Transparent user administration – single sign on (single grid identity), authorisation and accounting based on grid identity – AAA facilities
  - Job scheduling – which can handle different environments
  - Global data access
  - Global information services – job information, data information, resource information
- **Interoperability!**
  - Standards needed for federation of infrastructures – GGF, IETF....

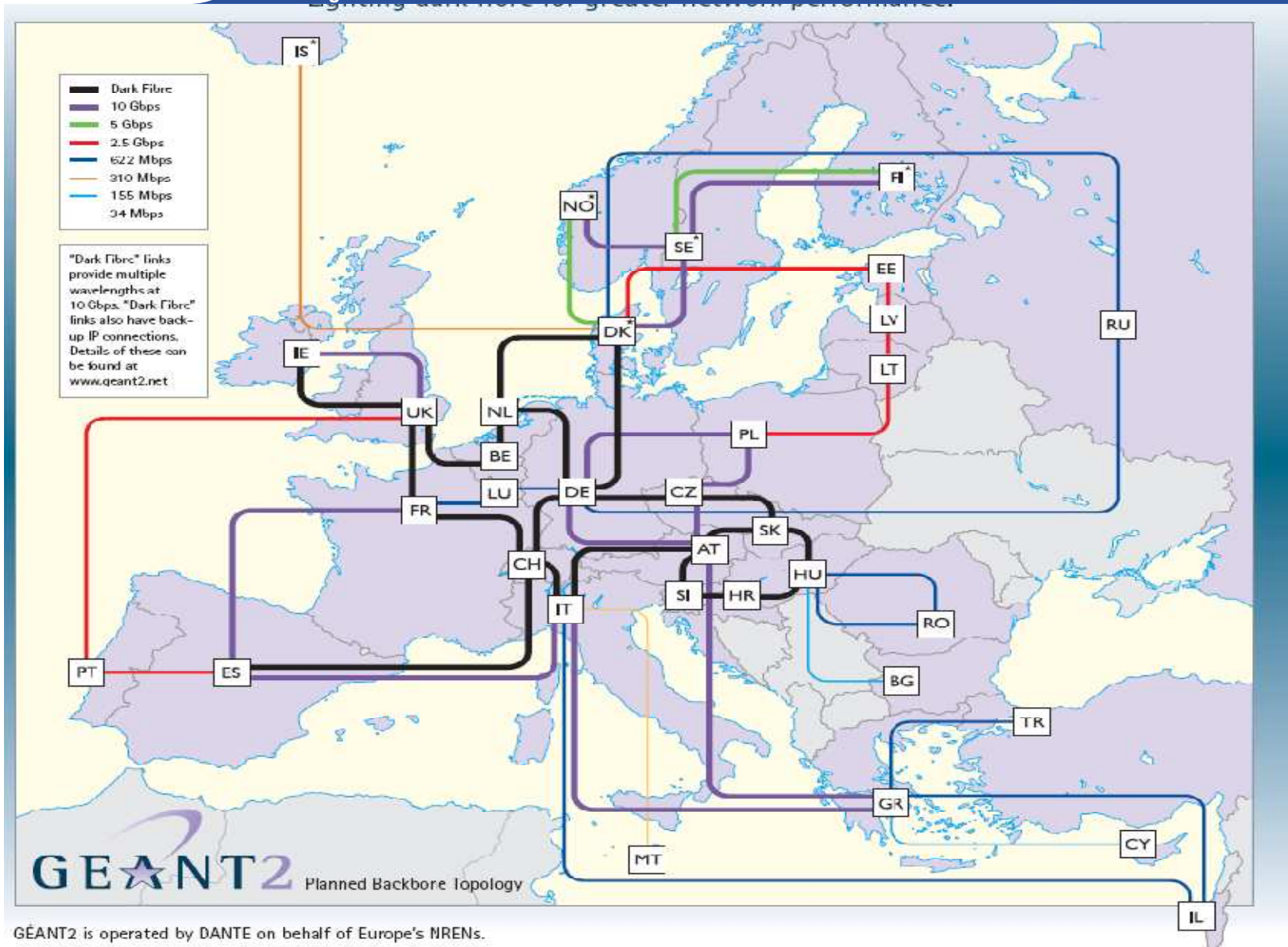


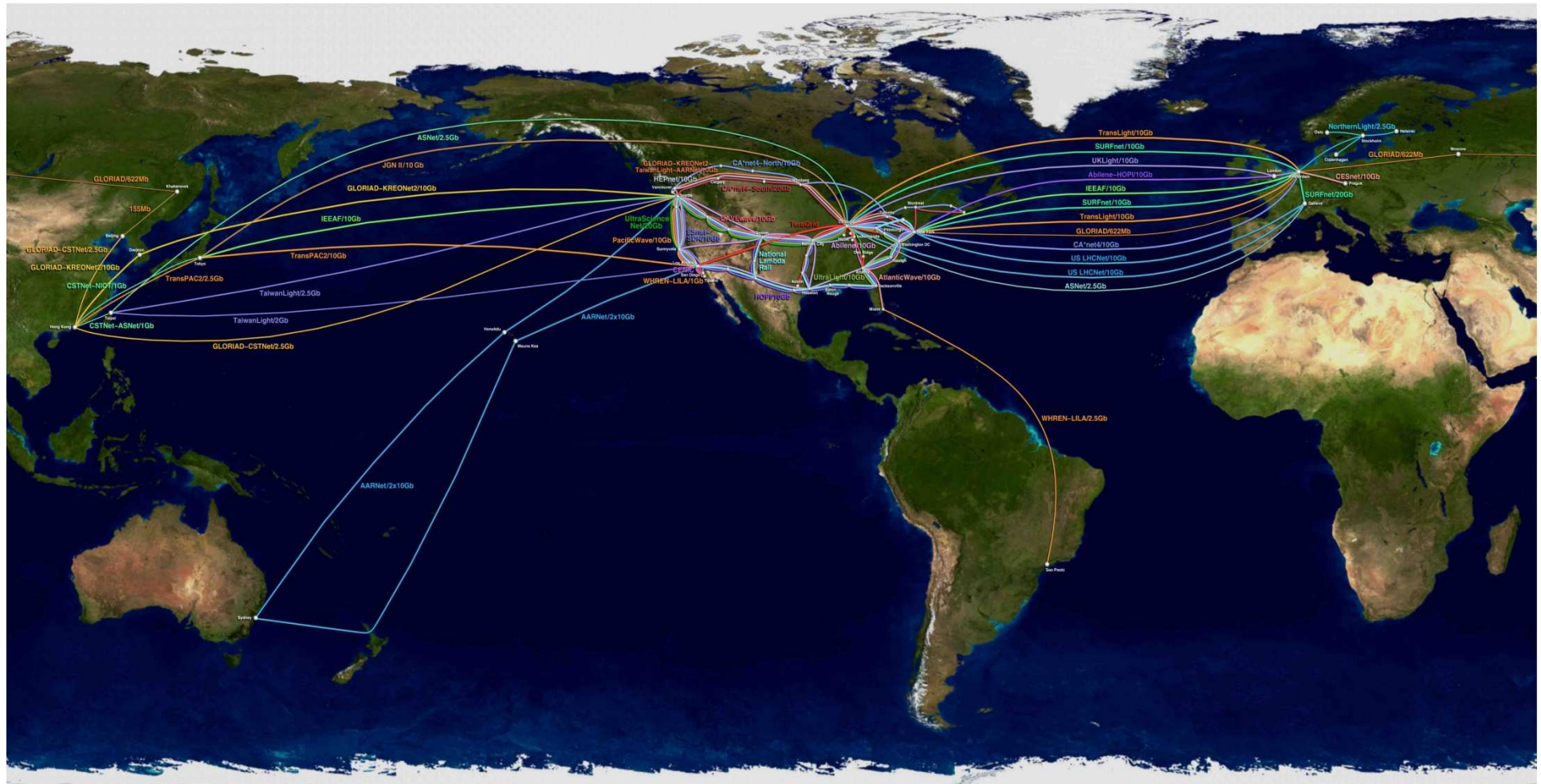
# It all starts with Networking

- **Developments in network connectivity (high bandwidths) and tools play an important role**
  - 10 Gbps WAN links available today, both shared links and dedicated lightpaths (based on lambda technology)
  - 1 Gbps network adapters are commodity items on systems today and 10GE adapters available



# GEANT2 topology





Visualization courtesy of Bob Patterson, NCSA/University of Illinois at Urbana-Champaign.

Data compilation by Maxine Brown, University of Illinois at Chicago. Earth texture from NASA.

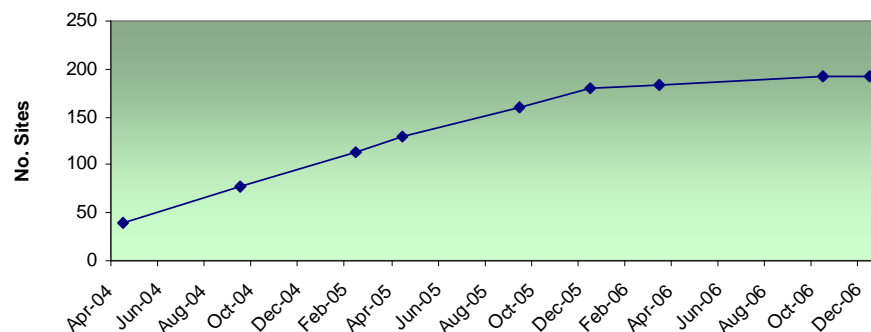
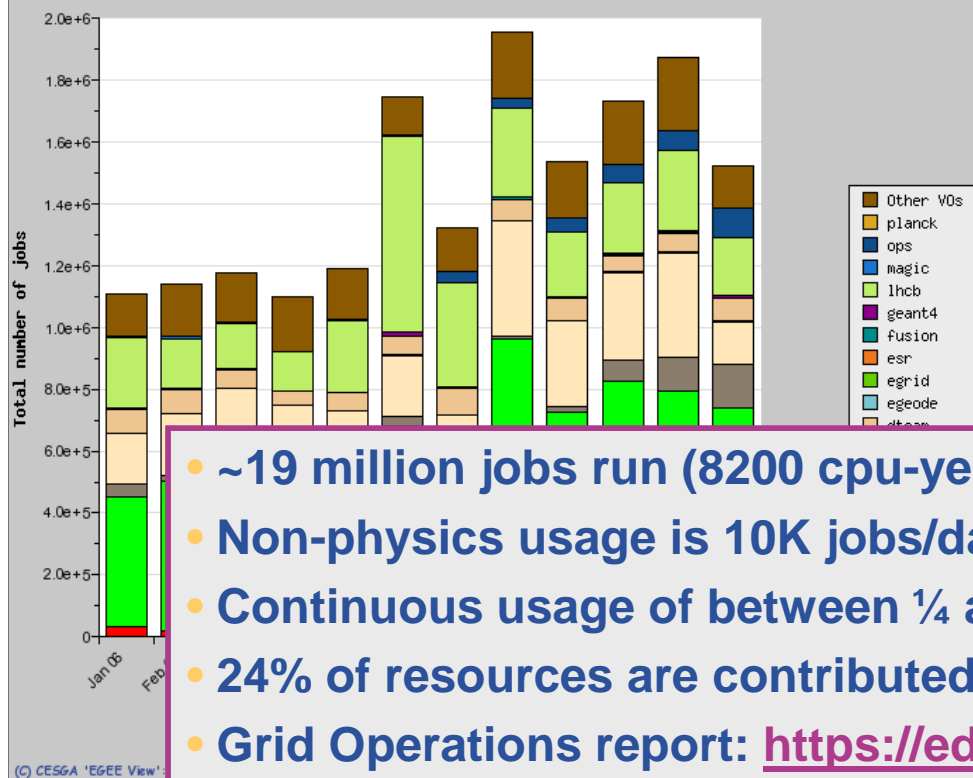


- **EGEE**
  - 1 April 2004 – 31 March 2006
  - 71 partners in 27 countries, federated in regional Grids
  - Operation of a pan European production Grid
- **EGEE-II**
  - 1 April 2006 – 31 March 2008
  - Expanded consortium
    - 91 partners
    - 11 Joint Research Units
  - Natural continuation of EGEE
  - Emphasis on providing production-level infrastructure
    - increased support for applications
    - interoperation with other Grid infrastructures
    - more involvement from Industry

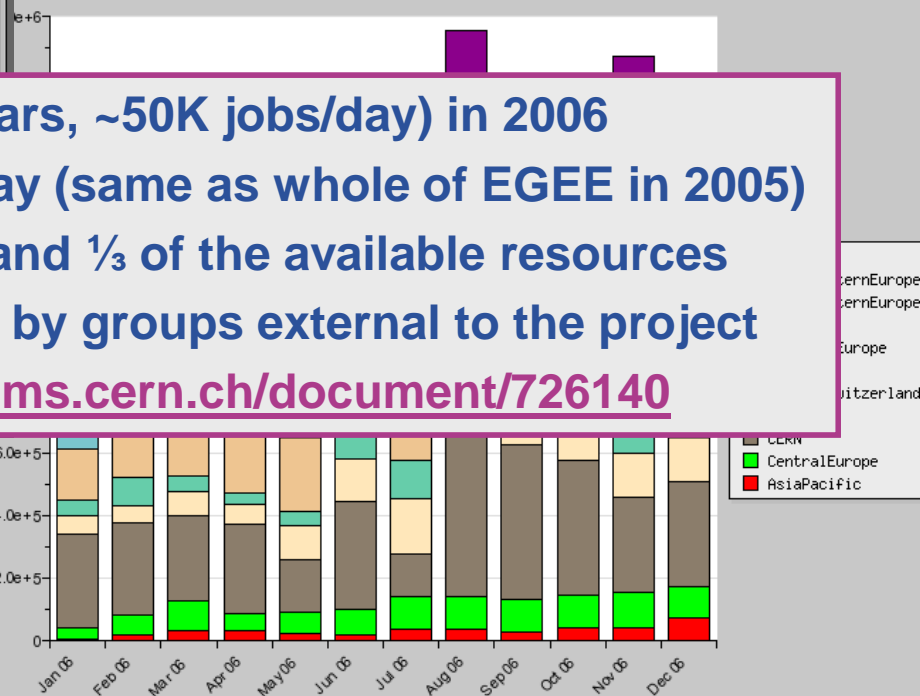
- **Mission**
  - Manage and operate production e-Infrastructure open to all user communities and service providers
  - Contribute to Grid standardisation and policy efforts
- **Infrastructure operation**
  - Currently include ~200 sites across 39 countries
  - Continuous monitoring of Grid services in a distributed global infrastructure
  - Automated site configuration/management
- **Future**
  - Expand on interoperability with related infrastructures

# Production Usage Status

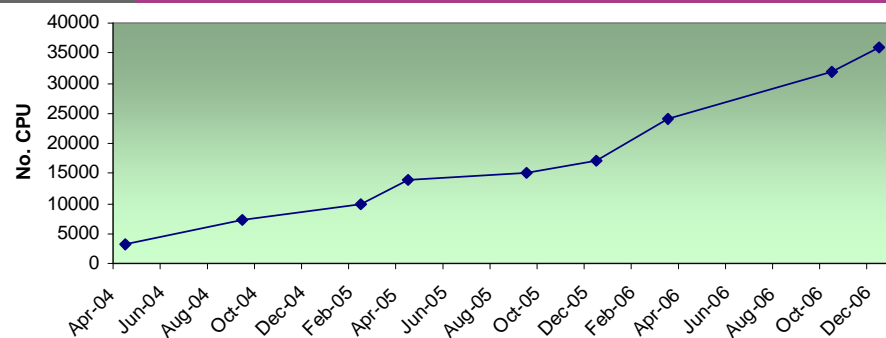
PRODUCTION Total number of jobs by VO and DATE  
January 2006 - December 2006

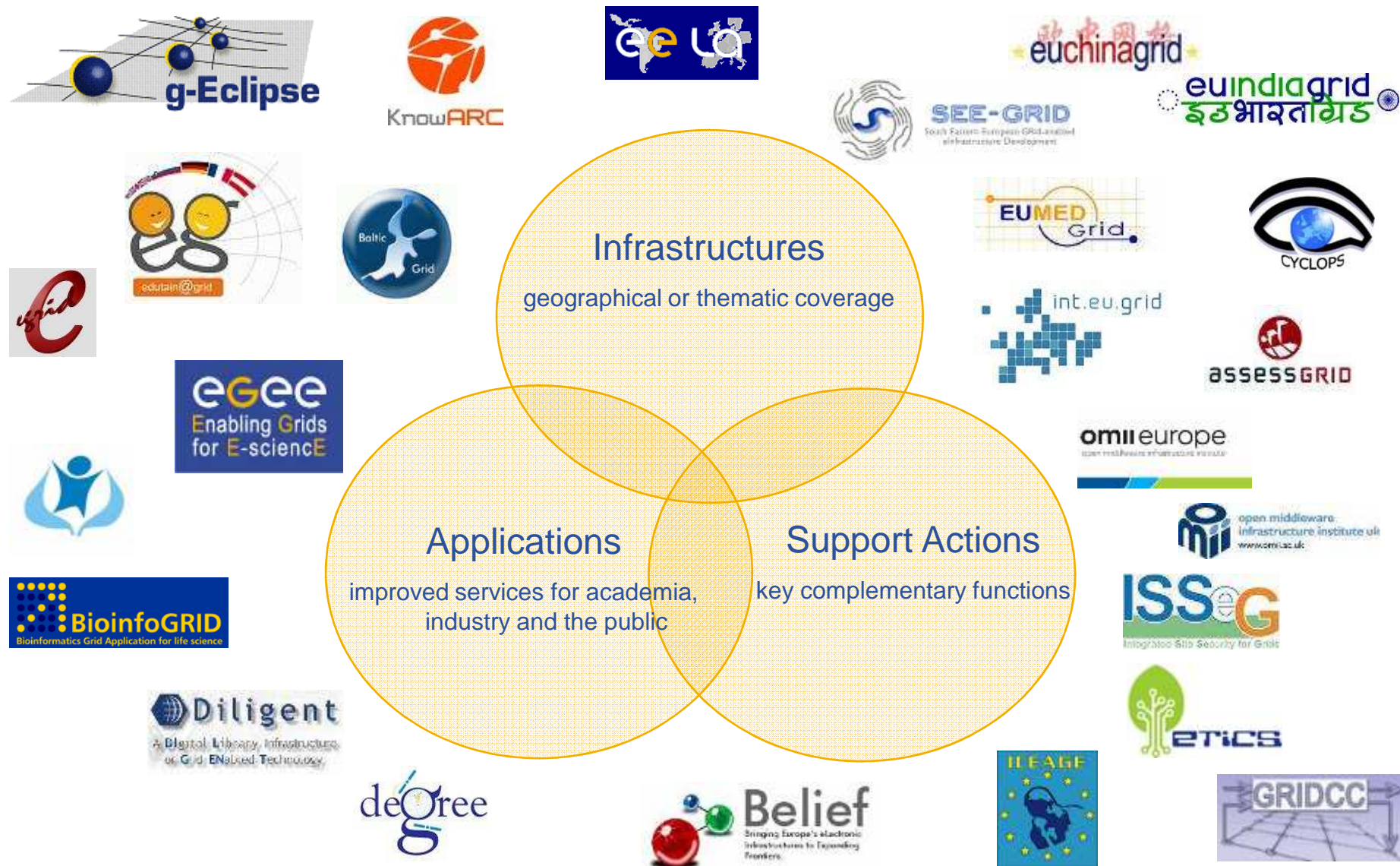


PRODUCTION Total number of jobs by REGION and DATE  
January 2006 - December 2006

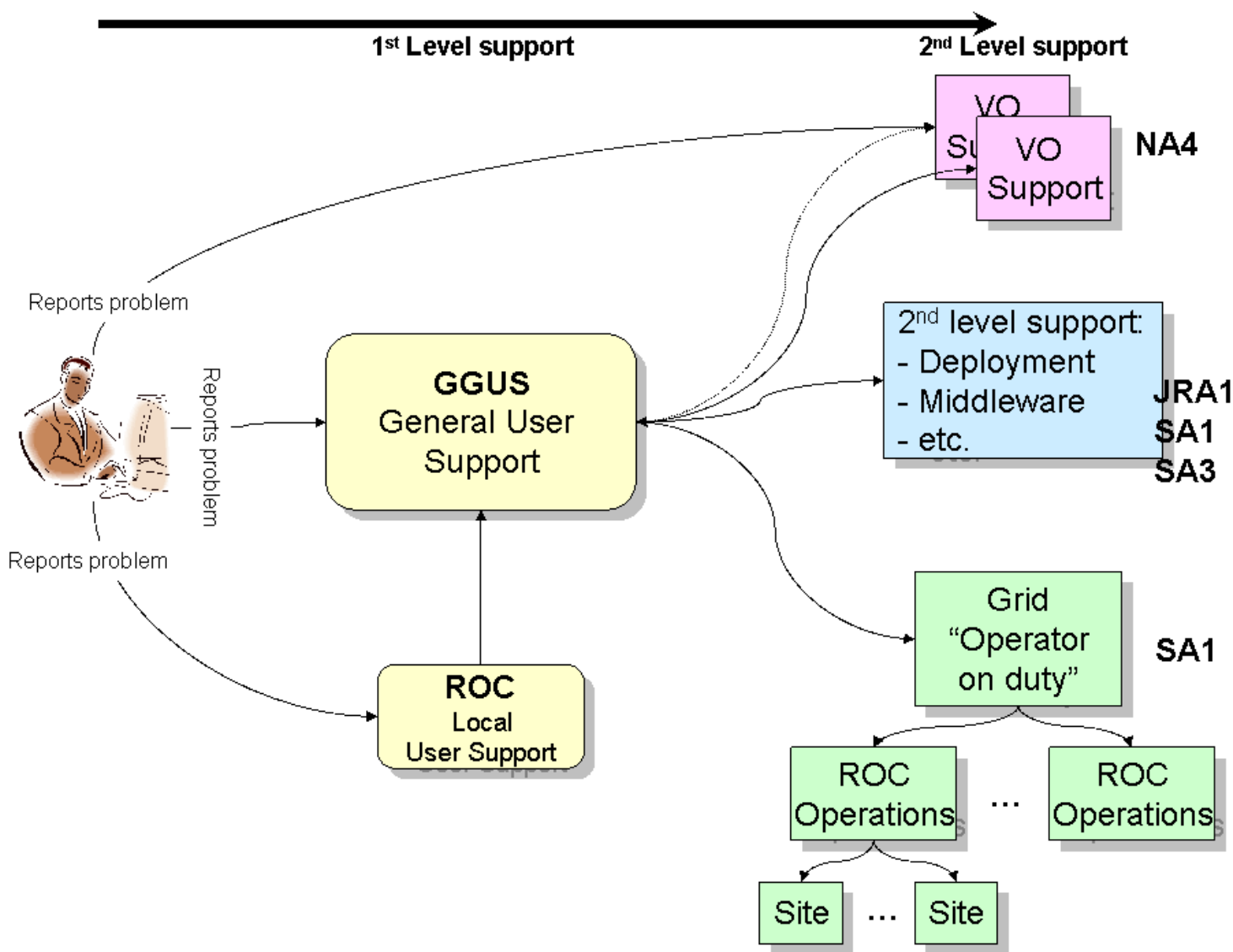


- ~19 million jobs run (8200 cpu-years, ~50K jobs/day) in 2006
- Non-physics usage is 10K jobs/day (same as whole of EGEE in 2005)
- Continuous usage of between  $\frac{1}{4}$  and  $\frac{1}{3}$  of the available resources
- 24% of resources are contributed by groups external to the project
- Grid Operations report: <https://edms.cern.ch/document/726140>





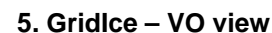




- **NE website:** <http://www.egee-ne.org/operations>
- **User support: contact user support at local site or mail to [support@egee-ne.org](mailto:support@egee-ne.org)**
  - NE uses a ticketing system monitored by different partners from our region. In NL NIKHEF, RC-RuG, SARA responsible.
  - Tickets from GGUS are also imported in the NE system
- **Application support – NA4 activity. In NL RC-RuG, SARA**



State	Site Name	URL	Test Date	Version	Software Version	AS/WHO Version	AS/WHO Status	AS/WHO Date	AS/WHO Comment	AS/WHO URL
China	Shanghai	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
China	BEIJING-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
South Korea	SEOUL-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Finland	HELSINKI-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Belgium	BRUSSELS-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
USA	NEW YORK-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Canada	OTTAWA-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
UK	LONDON-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Germany	FRANKFURT-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
France	PARIS-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Italy	MILAN-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Spain	MADRID-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Japan	TOKYO-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com
Australia	MELBOURNE-EDGE	api.cloudflare.com	2008-01-16	1.0.0	1.0.0	1.0.0	OK	2008-01-16	OK	http://api.cloudflare.com





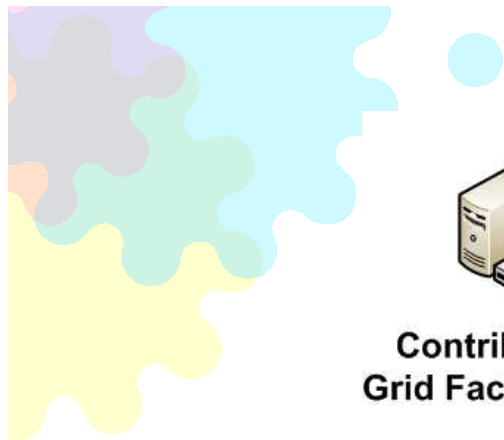


# BiG Grid

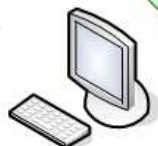
- **Strengthen existing National Grid infrastructure in Netherlands (NL-GRID by NCF)**
- **Sudsidy of 28 M€ for hardware and peopleware (expertise and support)**
- **Core partners**
  - NCF
  - Nikhef (High Energy Physics)
  - NBIC (BioInformaitcs)
- **Central and Distributed facilities**



**BiG Grid**  
the dutch e-science grid

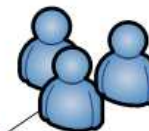


**Contributed  
Grid Facilities**



### Distributed Clusters

Cluster computing (grid accessible, complementary local use)  
Cache storage (limited D1T0)  
Access points and grid interfaces



### Central Facilities

Capacity computing  
Capability computing  
Database services  
Access services  
Mass storage  
(D1T0, D1T1, D0T1)



SARA



NIKHEF



RC-RUG



PHICOS\*



GEANT  
OPN  
Netherlight  
...

**SURFnet network**

**Contributed  
Grid Facilities**

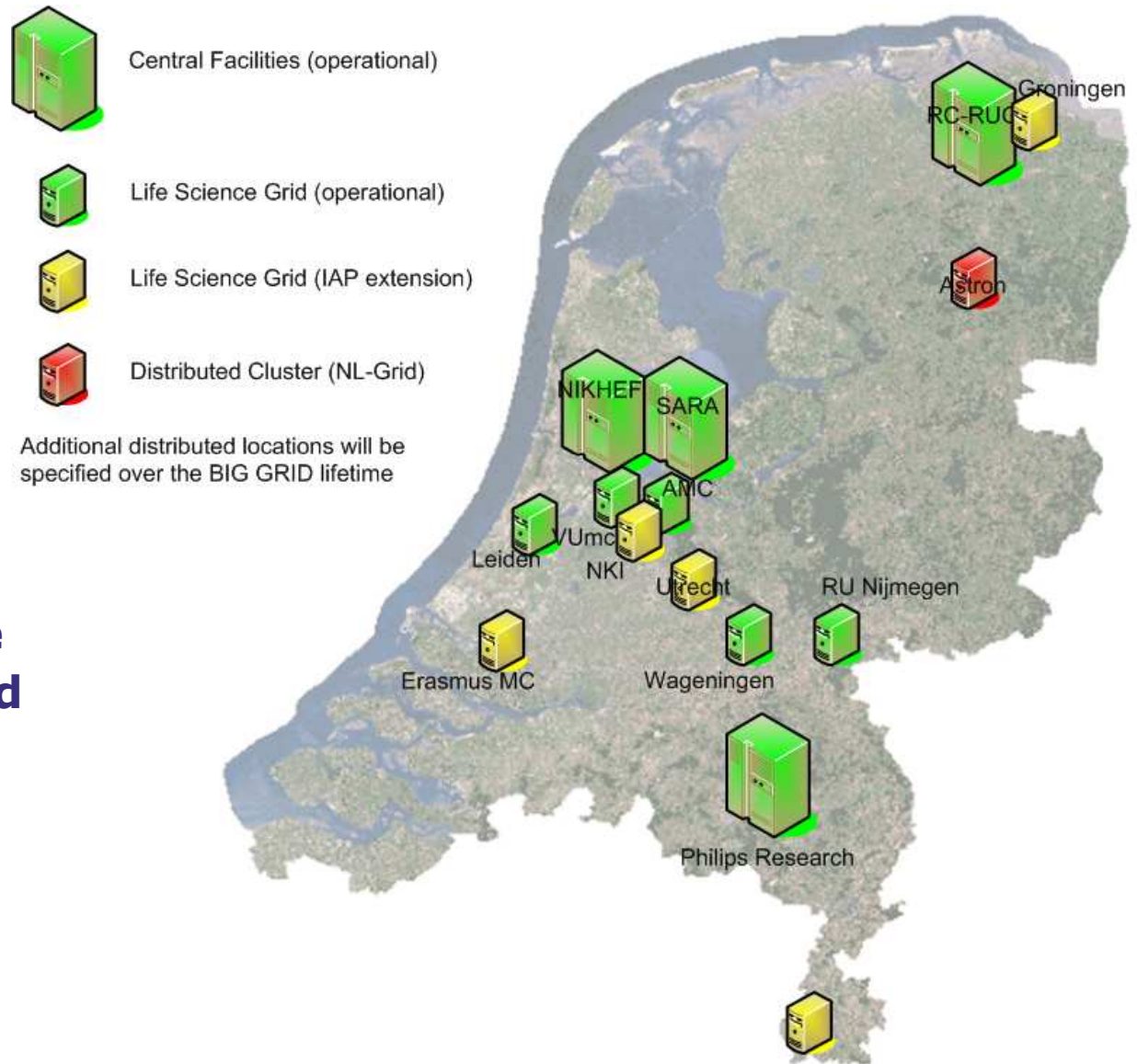


**BiG Grid**  
the dutch e-science grid



# Infrastructure

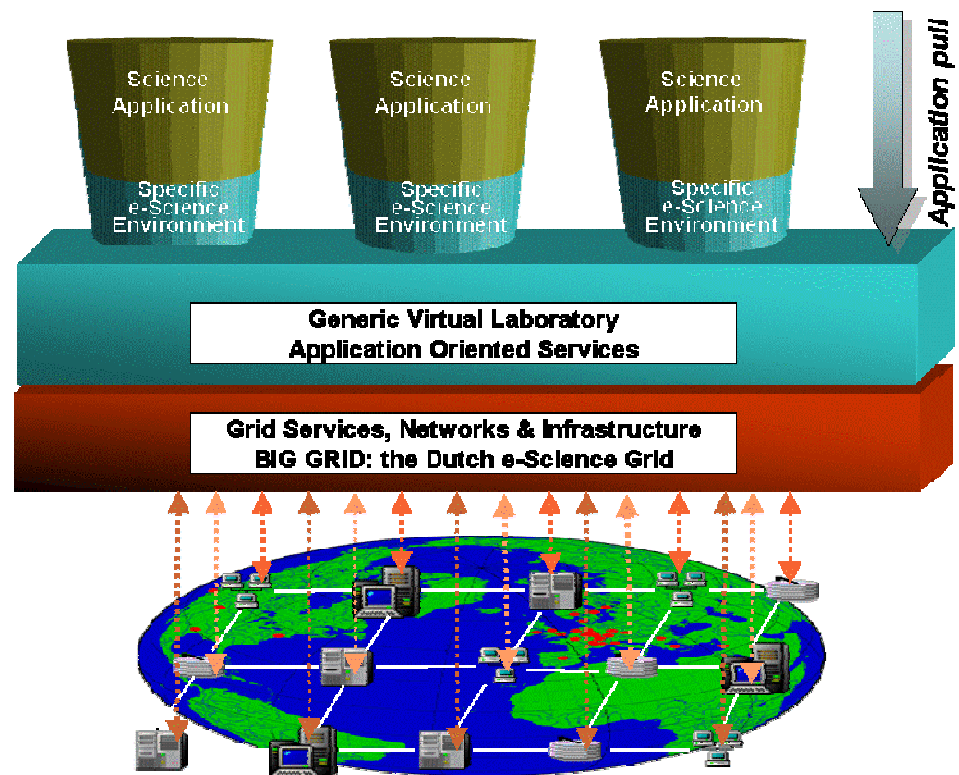
**O(5000) CPU**  
**O(10) PB disk storage**  
**O(20) PB tape storage**  
**O(10) Life Science Grid clusters**



**BiG Grid**  
the dutch e-science grid

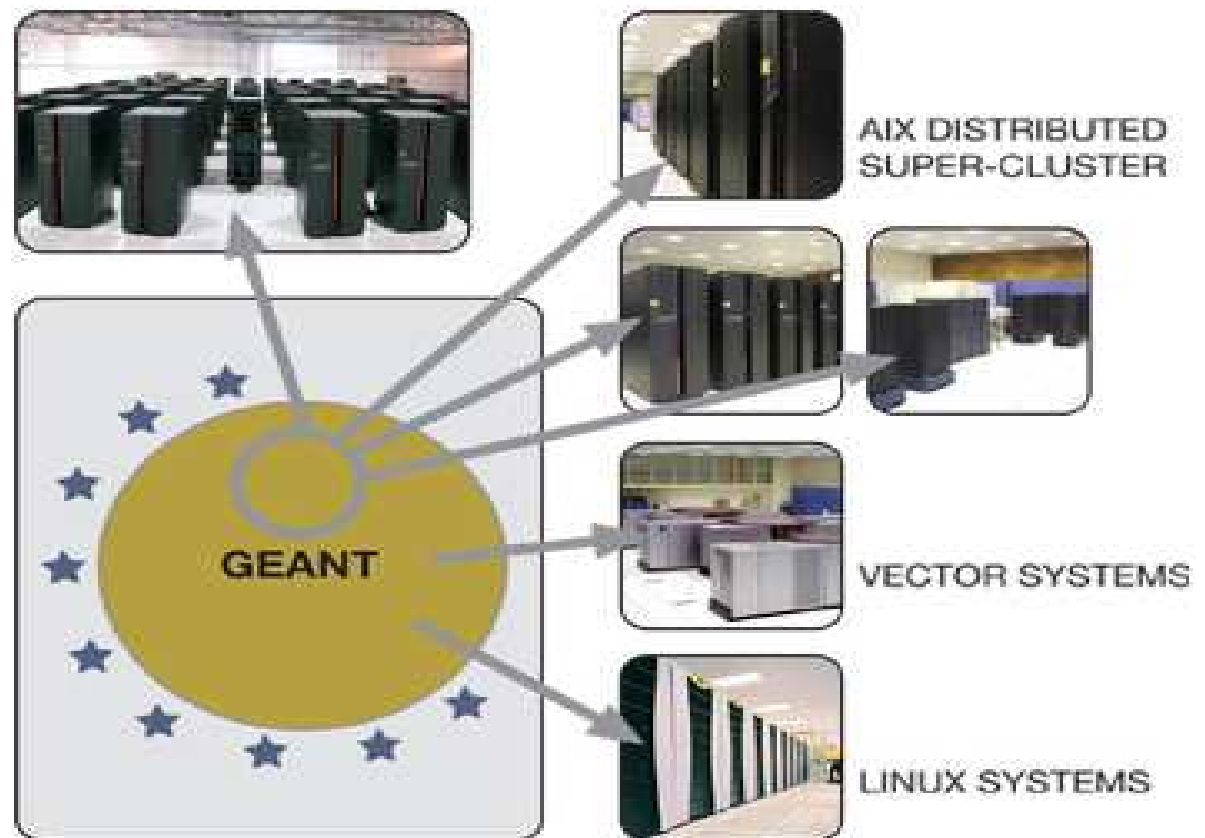
# Combination of 'push' and 'pull'

- **Application support:**
  - expertise: Application Domain Analysts
  - help desk and operations centre
- **Uniform software suite → Collaboration**
  - Standards:
    - Open Grid Forum
  - EGEE: 'production' grid (40 disciplines)
  - BSIK VL-E project





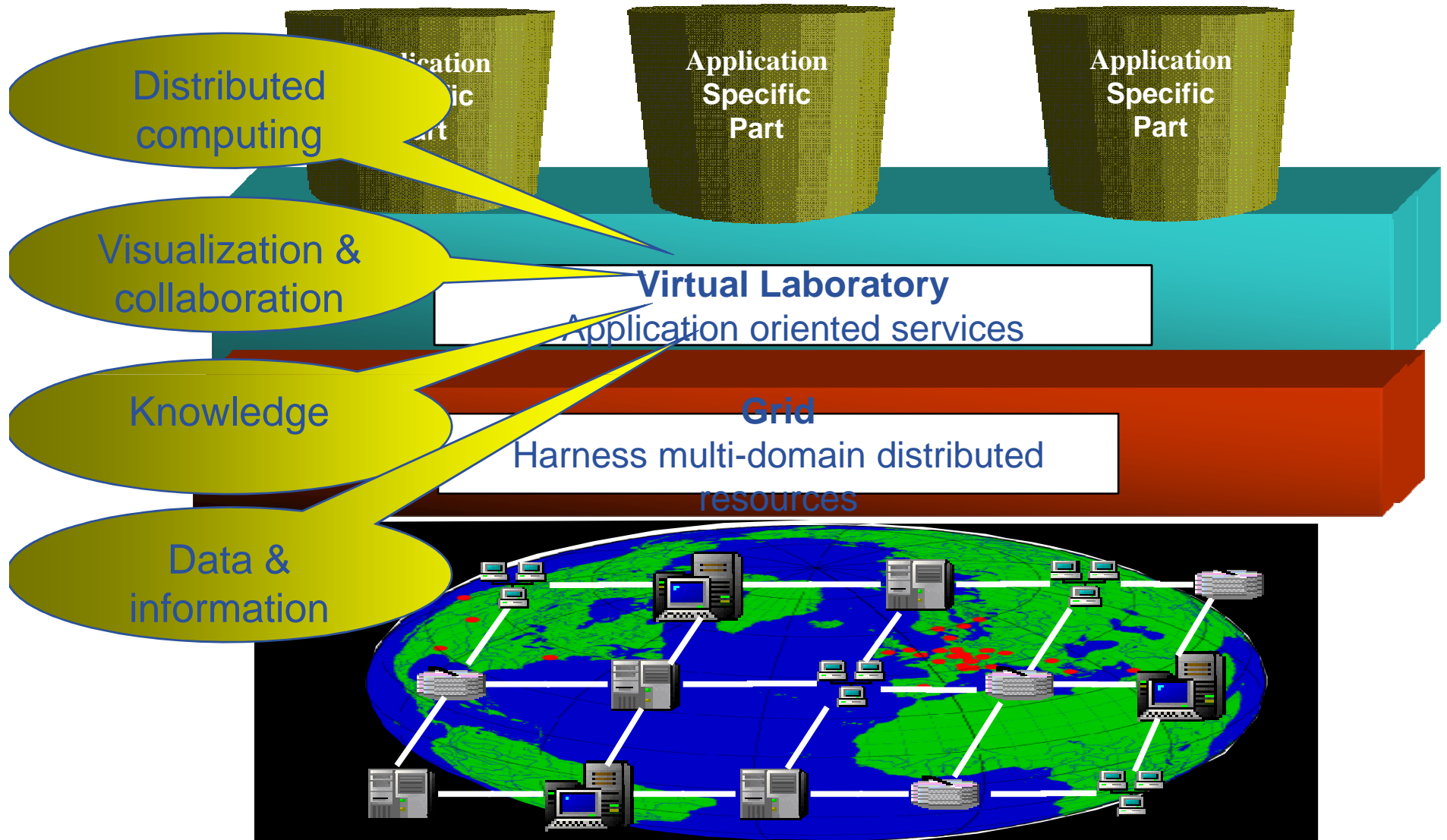
- European super-computing grid
- Shared global file system
- Job migration
- Co-scheduling



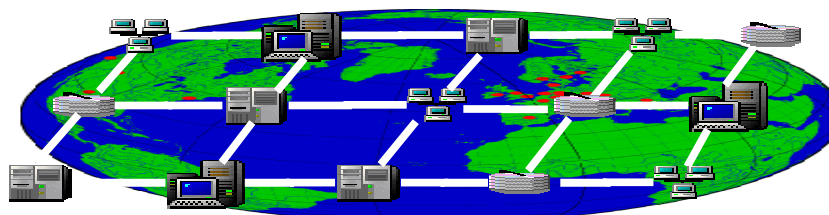
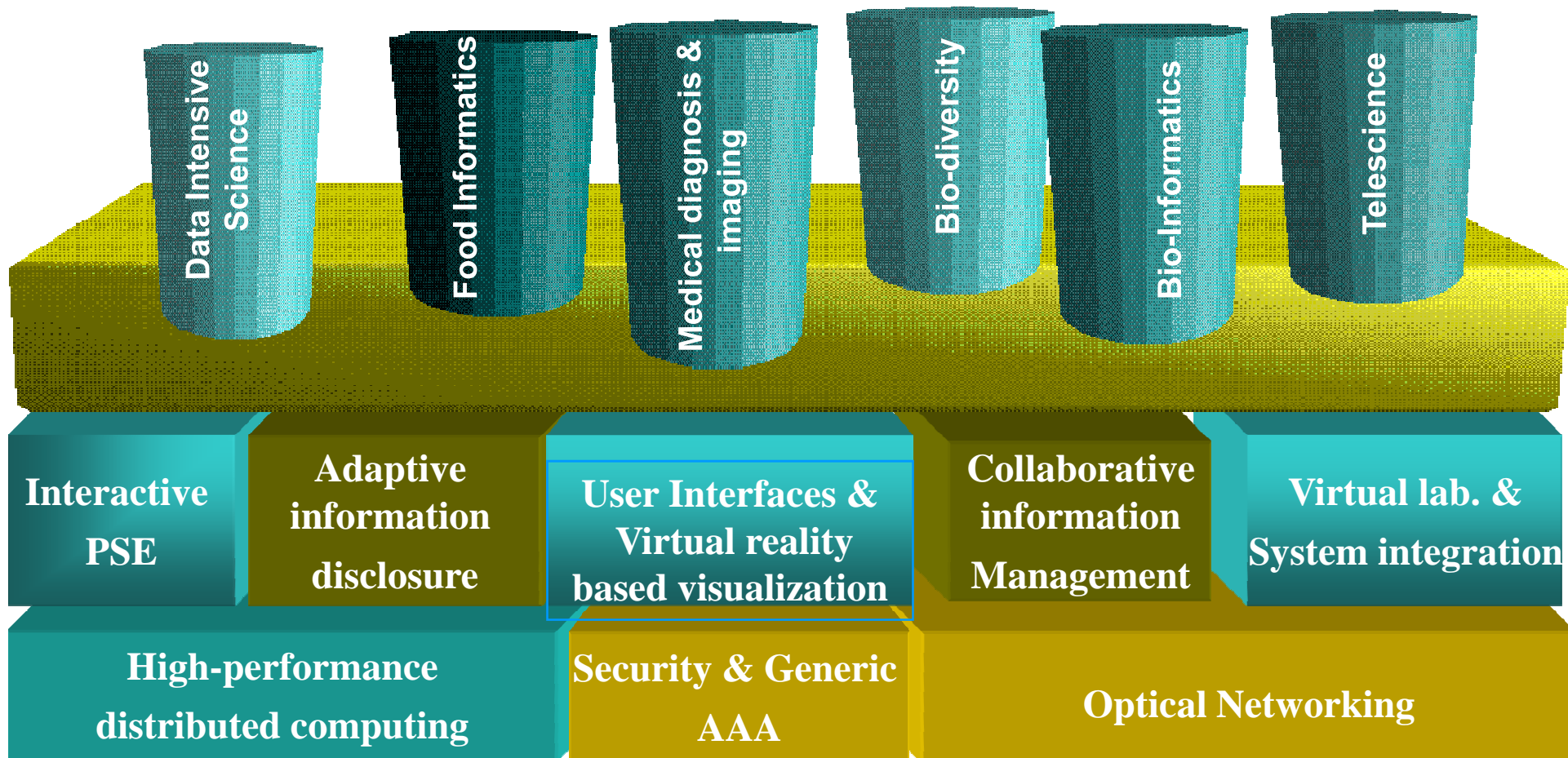
# DEISA

DISTRIBUTED EUROPEAN INFRASTRUCTURE FOR SUPERCOMPUTING APPLICATIONS

# Virtual Laboratories



# Virtual Laboratory for e-Science



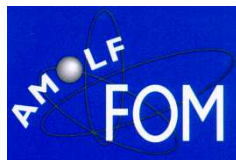
- 40 M€ (20 M€ BSIK funding)
- 2004 - 2008
- 20 partners
- Academic - Industrial



UNIVERSITEIT VAN AMSTERDAM



*vrije Universiteit*



- **Day 1**
  - Authentication
  - Job Submission
  - Tools for Collaboration by SURFnet
  - Drinks
  
- **Day 2**
  - Data handling
  - User Scenarios (VI-e and DANS)

Tutorial is free thanks to support of our sponsors



- Gridforum.nl



- Netherlands Center for Bioinformatics



- NCF



- SURFnet