vl-e virtual laboratory for e-science

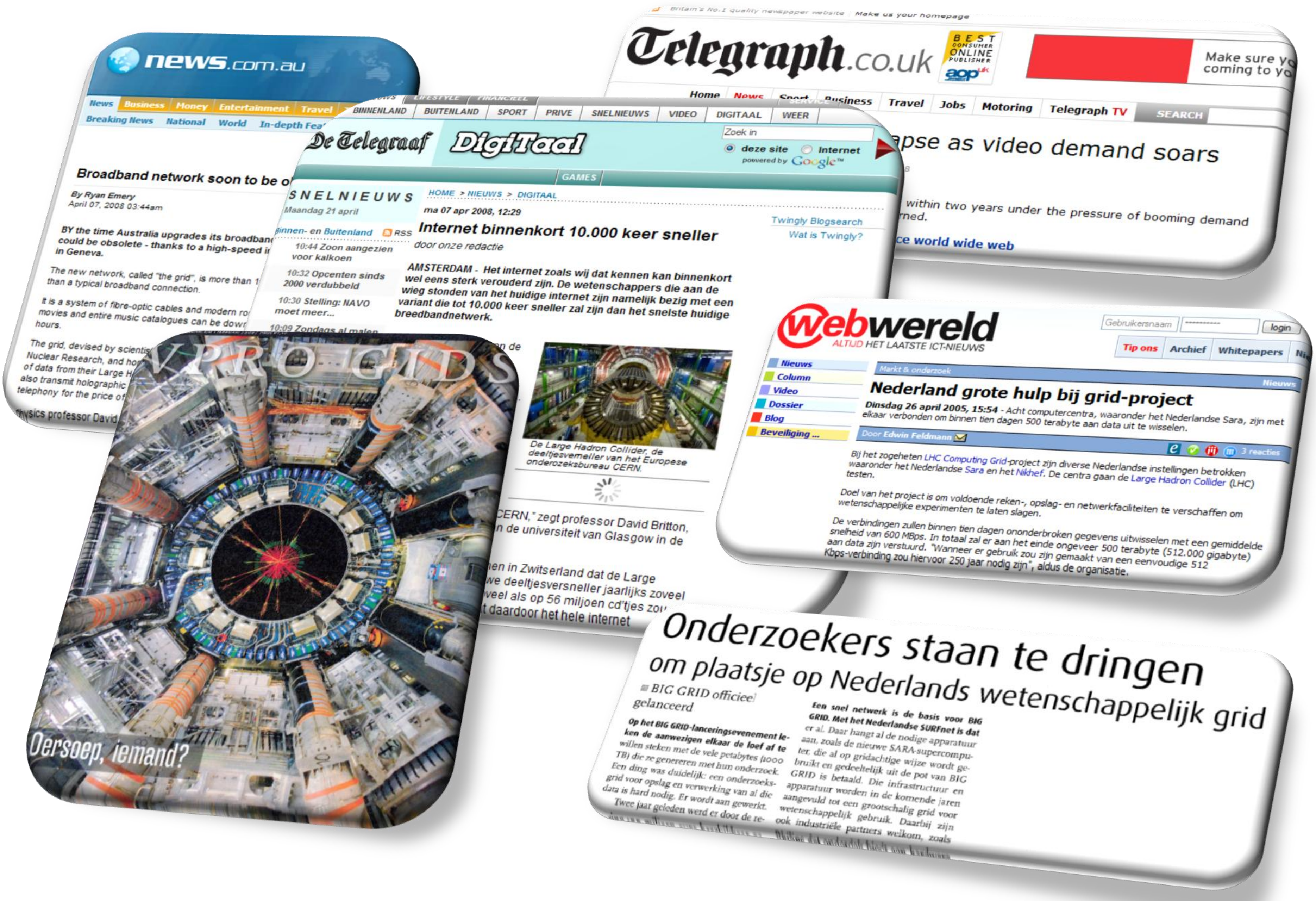
BiG Grid
the dutch e-science grid

Grid: data delen op wereldschaal

David Groep, Nikhef

Rotary Krommenie-Wormerveer
9 mei 2008

Graphics: Real Time Monitor
Gidon Moont, Imperial College London, see http://gridportal.hep.ph.ic.ac.uk/rtm/



news.com.au

News | Business | Money | Entertainment | Travel

Breaking News | National | World | In-depth Fea

Broadband network soon to be o

By Ryan Emery
April 07, 2008 03:44am

BY the time Australia upgrades its broadban could be obsolete - thanks to a high-speed i in Geneva.

The new network, called "the grid", is more than 1 than a typical broadband connection.

It is a system of fibre-optic cables and modern ro movies and entire music catalogues can be dowr hours.

The grid, devised by scientis Nuclear Research, and ho of data from their Large H also transmit holographic telephony for the price of

hysics professor David

Britain's No.1 quality newspaper website | Make us your homepage

Telegraph.co.uk

BEST CONSUMER ONLINE PUBLISHER aop uk

Make sure y coming to yo

Home | News | Sport | Business | Travel | Jobs | Motoring | Telegraph TV | SEARCH

apse as video demand soars

within two years under the pressure of booming demand rned.

ce world wide web

LIFESTYLE | FINANCIEEL

BINNENLAND | BUITENLAND | SPORT | PRIVE | SNELNIEUWS | VIDEO | DIGITAAL | WEER

De Telegraaf DigiTaal

Zoek in
● deze site ○ Internet
powered by Google™

GAMES

SNELNIEUWS
Maandag 21 april

Binnen- en Buitenland RSS

10:44 Zoon aangezien voor kalkoen

10:32 Opcenten sinds 2000 verdubbeld

10:30 Stelling: NAVO moet meer...

10:09 Zondags al malen

HOME > NIEUWS > DIGITAAL

ma 07 apr 2008, 12:29

Internet binnenkort 10.000 keer sneller

door onze redactie

Twingly Blogsearch
Wat is Twingly?

AMSTERDAM - Het internet zoals wij dat kennen kan binnenkort wel eens sterk verouderd zijn. De wetenschappers die aan de wieg stonden van het huidige internet zijn namelijk bezig met een variant die tot 10.000 keer sneller zal zijn dan het snelste huidige breedbandnetwerk.

n de

De Large Hadron Collider, de deeltjesversneller van het Europese onderzoeksbureau CERN.

CERN," zegt professor David Britton, n de universiteit van Glasgow in de

en in Zwitserland dat de Large we deeltjesversneller jaarlijks zoveel veel als op 56 miljoen cd'tjes zou t daardoor het hele internet

VPRO GIDS

Oersoep, iemand?

webwereld
ALTIJD HET LAATSTE ICT-NIEUWS

Gebruikersnaam | login

Tip ons | Archief | Whitepapers | Ni

Nieuws
Column
Video
Dossier
Blog
Beveiliging ...

Markt & onderzoek | Nieuws

Nederland grote hulp bij grid-project

Dinsdag 26 april 2005, 15:54 - Acht computercentra, waaronder het Nederlandse Sara, zijn met elkaar verbonden om binnen tien dagen 500 terabyte aan data uit te wisselen.

Door Edwin Feldmann

3 reacties

Bij het zogeheten LHC Computing Grid-project zijn diverse Nederlandse instellingen betrokken waaronder het Nederlandse Sara en het Nikhef. De centra gaan de Large Hadron Collider (LHC) testen.

Doel van het project is om voldoende reken-, opslag- en netwerkfaciliteiten te verschaffen om wetenschappelijke experimenten te laten slagen.

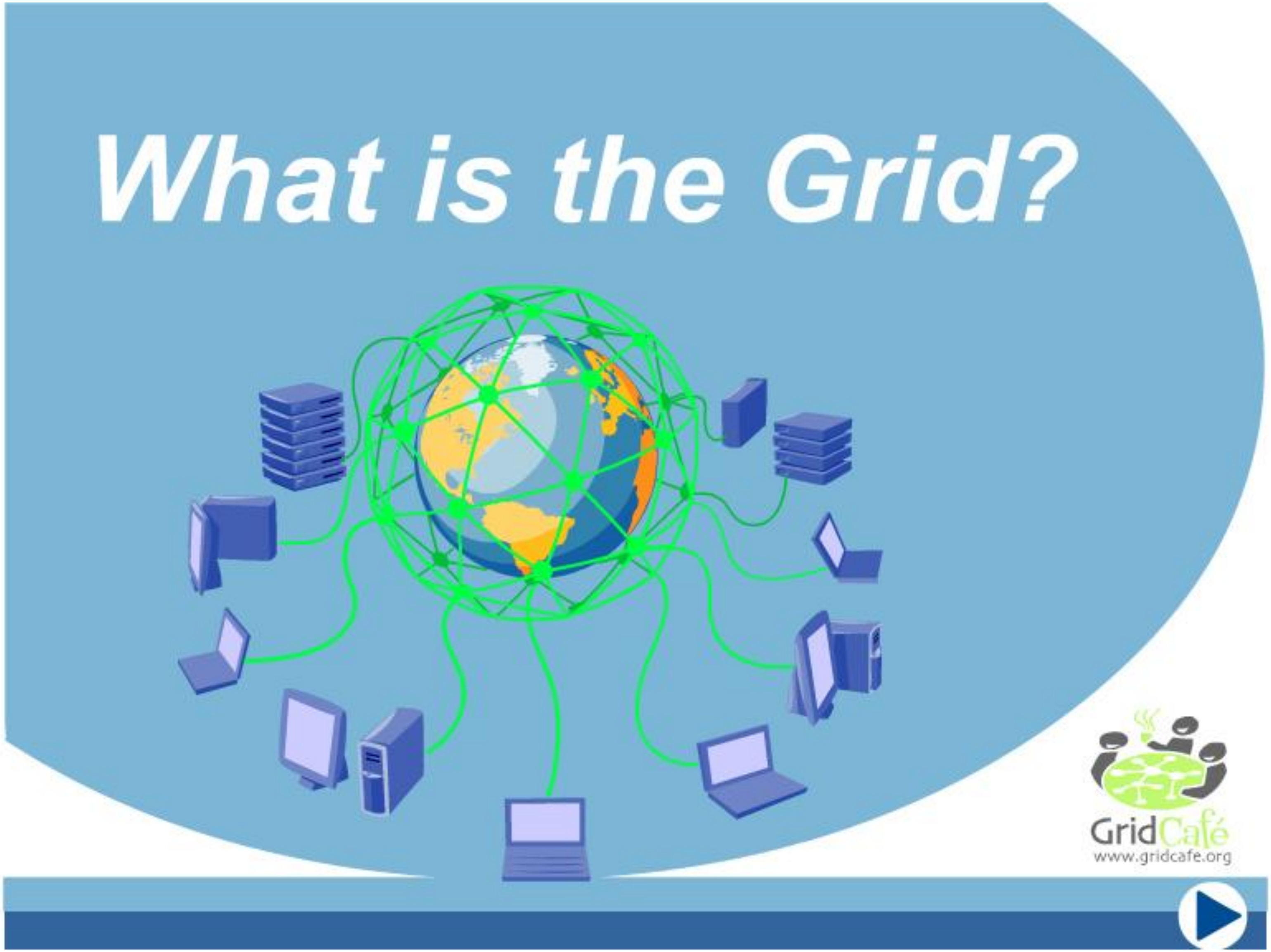
De verbindingen zullen binnen tien dagen ononderbroken gegevens uitwisselen met een gemiddelde snelheid van 600 MBps. In totaal zal er aan het einde ongeveer 500 terabyte (512.000 gigabyte) aan data zijn verstuurd. "Wanneer er gebruik zou zijn gemaakt van een eenvoudige 512 Kbps-verbinding zou hiervoor 250 jaar nodig zijn", aldus de organisatie.

Onderzoekers staan te dringen om plaatsje op Nederlands wetenschappelijk grid

BIG GRID officieel gelanceerd

Op het BIG GRID-lanceringsevenement leken de aanwezigen elkaar de loef af te willen steken met de vele petabytes (1000 TB) die ze genereren met hun onderzoek. Een ding was duidelijk: een onderzoeksgrid voor opslag en verwerking van al die data is hard nodig. Er wordt aan gewerkt. Twee jaar geleden werd er door de re-

Een snel netwerk is de basis voor BIG GRID. Met het Nederlandse SURFnet is dat er al. Daar hangt al de nodige apparatuur aan, zoals de nieuwe SARA-supercomputer, die al op gridachtige wijze wordt gebruikt en gedeeltelijk uit de pot van BIG GRID is betaald. Die infrastructuur en apparatuur worden in de komende jaren aangevuld tot een grootschalig grid voor wetenschappelijk gebruik. Daarbij zijn ook industriële partners welkom, zoals



What is the Grid?

Slide courtesy: GridCafe.org and Bob Jones, EGEE (www.eu-egee.org)

- Name “Grid” chosen by analogy with electric power grid (Foster and Kesselman 1997)
- Vision: plug-in computer for processing power just like plugging in toaster for electricity.

The idea has been around for decades

‘distributed computing’, *‘metacomputing’*

- *and will be around: ‘Web 2.0’, ‘Virtualisation’, ‘Cloud Computing’*

the Grid vision is to realise this on a global scale

Grids in Science

The Grid is 'more of everything' as science struggles to deal with ever increasing complexity

more than one place on earth

Real Time Monitor 2D (applet)

more than one computer

more than one science!

more than …

Why would we need it?

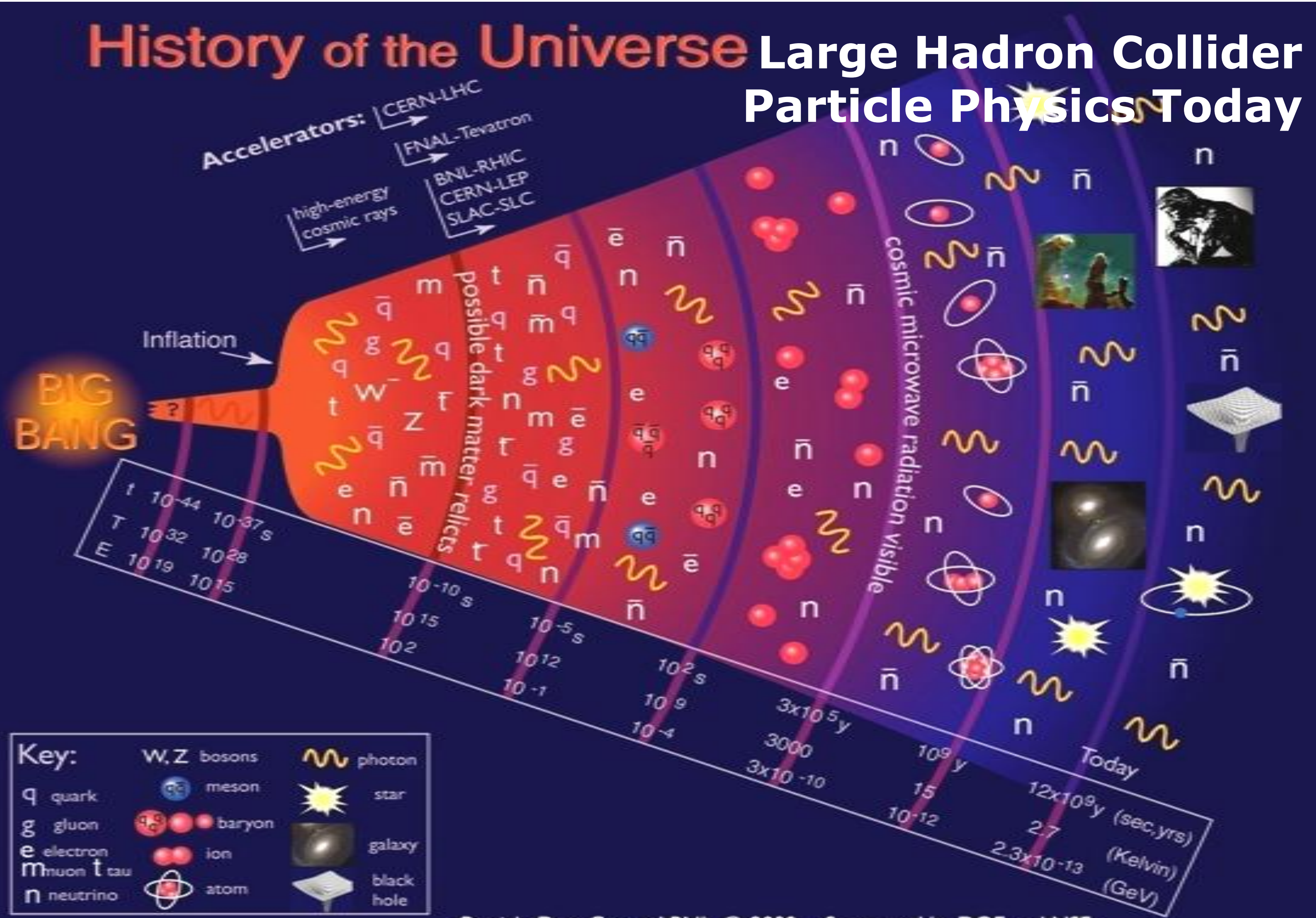
Enhanced Science needs more and more computations and

Collected data in science and industry grows exponentially

The Bible	**5 MByte**
X-ray image	5 MByte/image
Functional MRI	1 GByte/day
Bio-informatics databases	500 GByte each
Refereed journal papers	1 TByte/yr
Satellite world imagery	5 TByte/yr
US LoC contents	20 TByte
Internet Archive 1996-2002	100 TByte
Particle Physics 2005	1 PByte/yr
Particle Physics Today: LHC	**20 PByte/yr**

1 Petabyte = 1 000 000 000 Megabyte

Large Hadron Collider Particle Physics Today

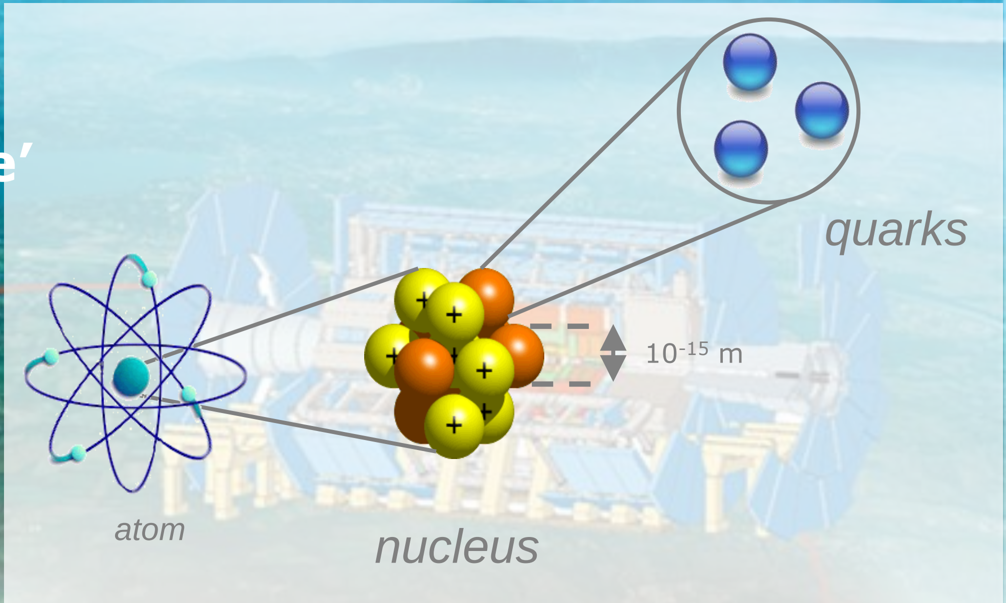

History of the Universe
Accelerators:
CERN-LHC
FNAL-Tevatron
BNL-RHIC
CERN-LEP
SLAC-SLC
high-energy cosmic rays
Inflation
BIG BANG
?
possible dark matter rellicts
cosmic microwave radiation visible
t 10-44 10-37 s 10-10 s 10-5 s 102 s 3x105 y 109 y Today 12x109 y (sec,yrs)
T 1032 1028 1015 1012 109 3000 15 2.7 (Kelvin)
E 1019 1015 102 10-1 10-4 3x10-10 10-12 2.3x10-13 (GeV)
Key:
q quark
g gluon
e electron
m muon t tau
n neutrino
W, Z bosons
meson
baryon
ion
atom
photon
star
galaxy
black hole

Particle Data Group, LBNL, © 2000. Supported by DOE and NSF

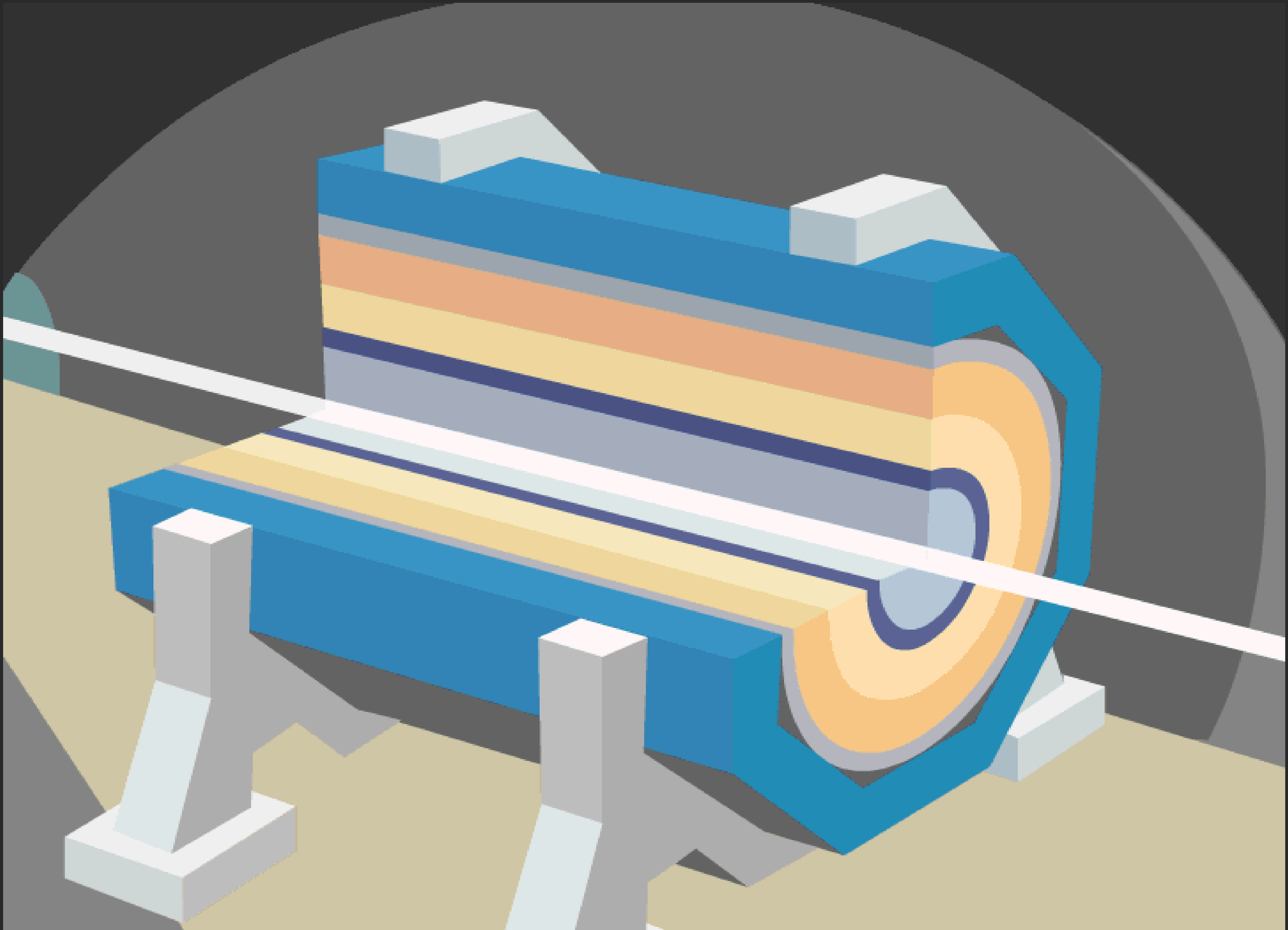
LHC Computing

Large Hadron Collider

- ‘the worlds largest microscope’
- ‘looking at the fundamental forces of nature’
- 27 km circumference
- Located at CERN, Geneva, CH



~ 20 PByte of data per year, ~ 60 000 modern PC style computers



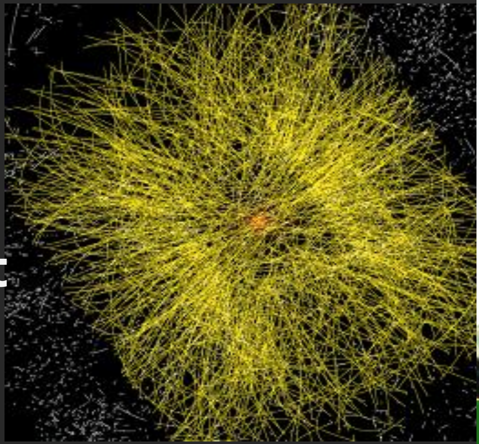
- Signal/Background 10^{-9}

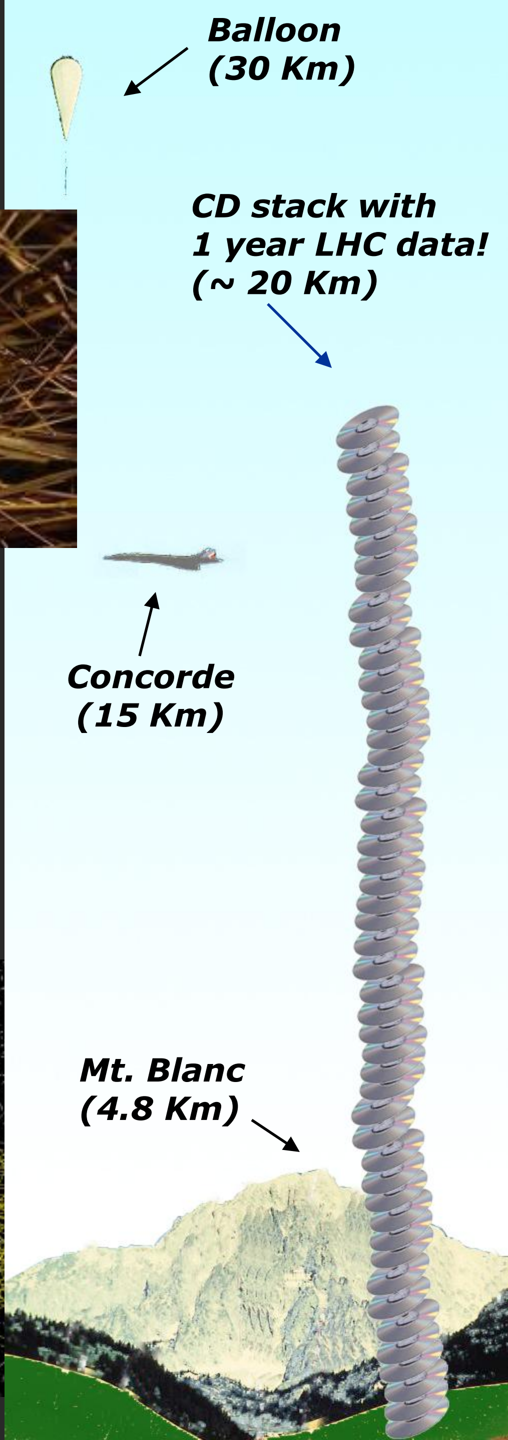
- Data volume
  - (high rate) **X**
    (large number of channels) **X**
    (4 experiments)

  ➔ **20 PetaBytes of new data each year**

- Compute power
  - (event complexity) **X**
    (number of events) **X**
    (thousands of users)

  ➔ **60'000 of (today's) fastest CPUs**





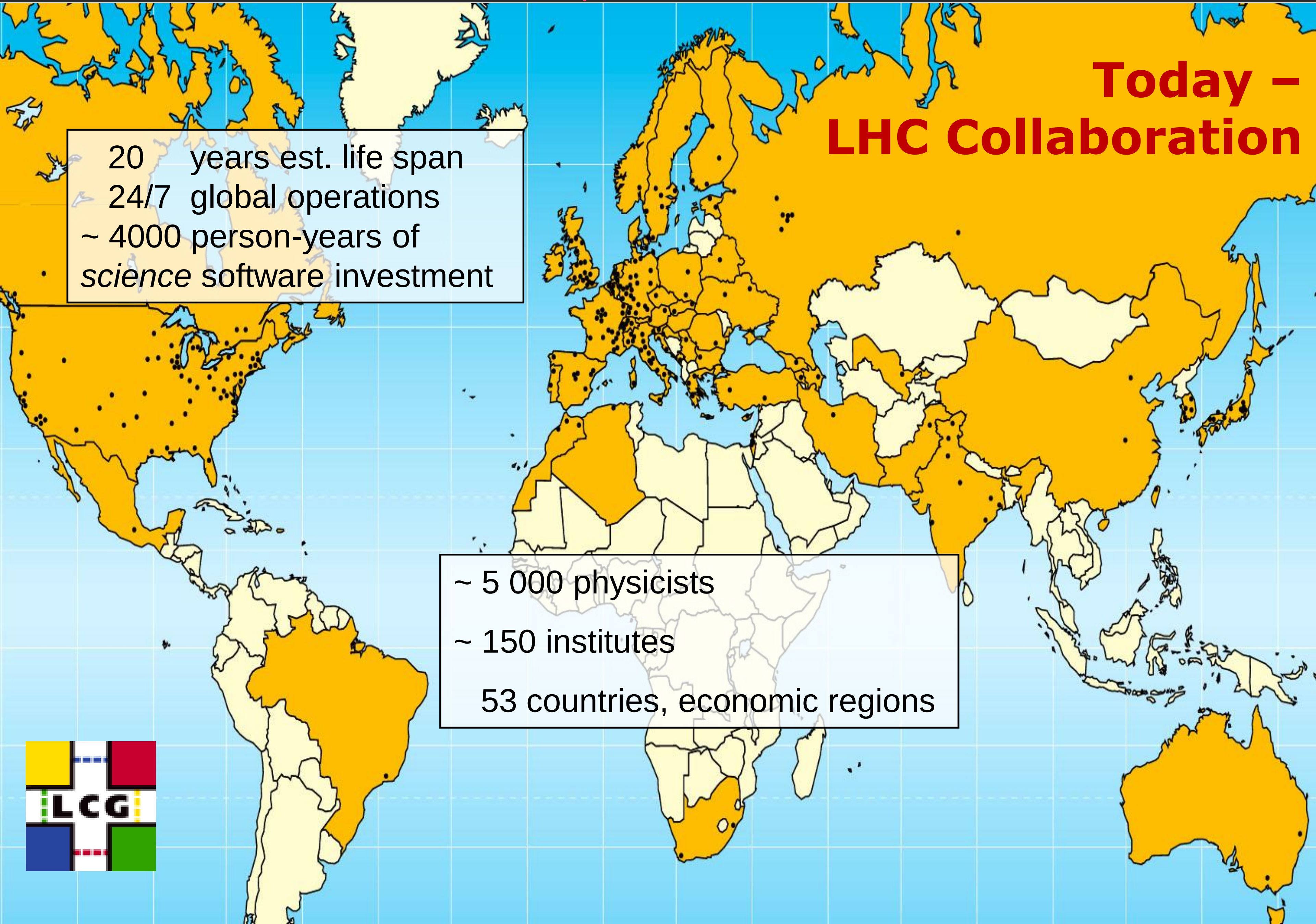
CERN, Where the web was born …

- Previous generation of HEP experiments (LEP) involved hundreds of scientists, thousands of engineers, and people working remotely

- Users at CERN, founded 1954 as Europe's first international organisation,
needed worldwide information sharing

- This need to share information inspired Tim Berners-Lee to create the 'World Wide Web' in 1990

Today – LHC Collaboration



20 years est. life span
24/7 global operations
~ 4000 person-years of *science* software investment

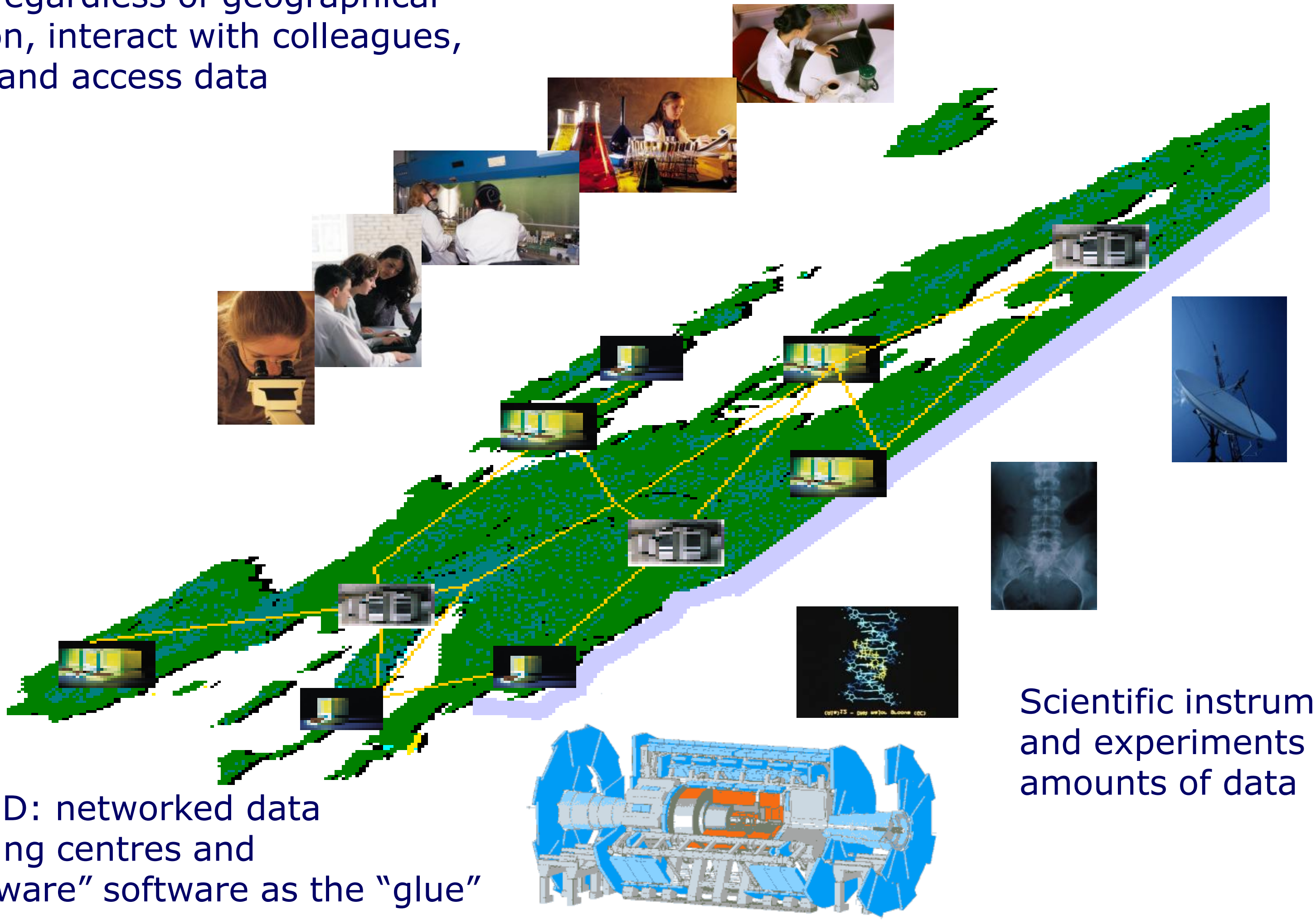
~ 5 000 physicists

~ 150 institutes

53 countries, economic regions

Beyond the Web: Grid for LHC and Science

Work regardless of geographical location, interact with colleagues, share and access data



The GRID: networked data processing centres and ”middleware” software as the “glue” of resources (computers, disks, mass storage).

Scientific instruments, libraries and experiments provide huge amounts of data

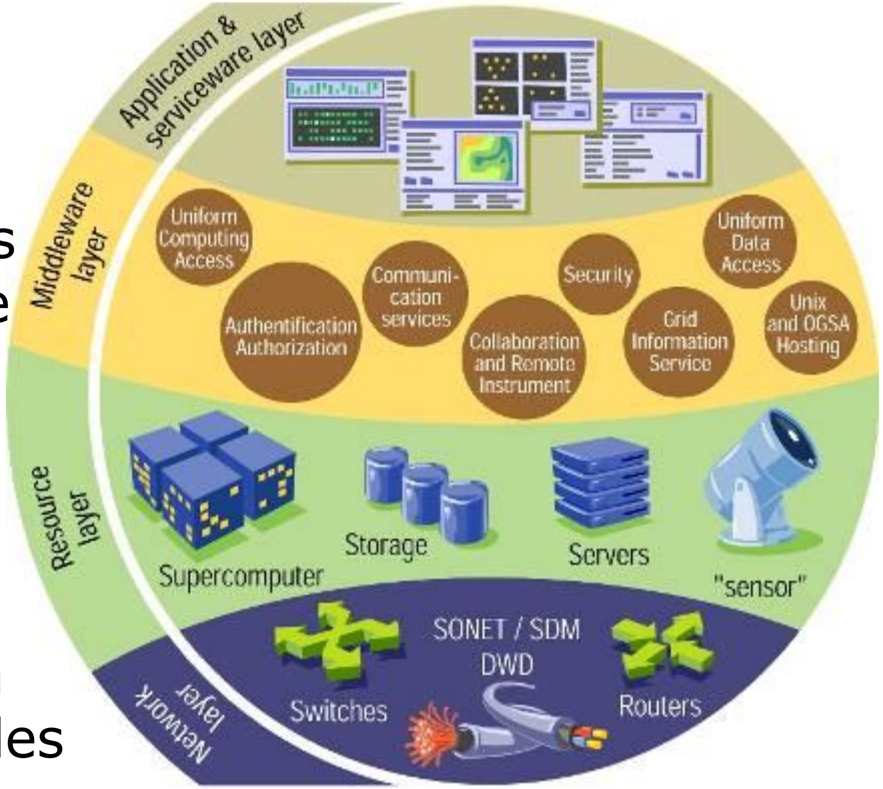
based on: Federico.Carminati@cern.ch

eGee
Enabling Grids for E-sciencE
Scheduled = 6849
Running = 10359
09:26:06 UTC
GridPP
UK Computing for Particle Physics

Making the Grid …

How does the Grid work?

- It relies on advanced software, called middleware.
- Middleware automatically locates data the scientist needs, and the computing power to analyse it.
- Middleware balances the load on different resources. It also handles security, accounting, monitoring and much more.

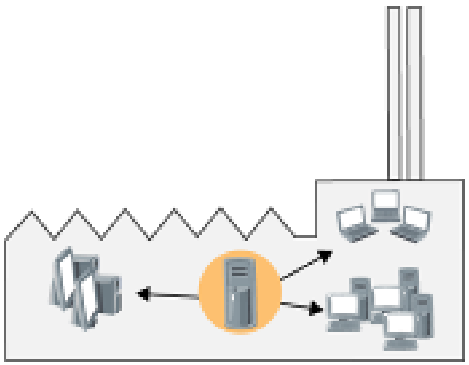


Slide courtesy: GridCafe.org and Bob Jones, EGEE (www.eu-egee.org)

Different Grids for different needs

- There is as yet no unified Grid (like there is a single web) rather there are many Grids for many applications.

- ‘Grid’ is used for different types of distributed computing
  - Enterprise Grids (within one company)
  - public resource Grids (volunteer your own PC).
  - scientific Grids that link together major computing centres in research labs and universities, who then federate to achieve *a global Grid infrastructure*



Corporate and commercial ‘Grids’

Large enterprises: finance, pharma, aerospace, cinema
… but …
some technologies based on grid concepts now offered as ‘hosted services’, also to SMEs

- ‘Backup as a Service’
  - *commercially available in NL*
- ‘Software as a Service’
  - *getting there bit by bit,*
    e.g. Google Apps, administrative software
- …

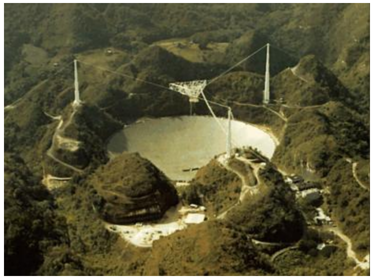
But ‘last mile’ network limitations in homes and SMEs limit more wide-spread use of grid technologies today

Contributed 'Volunteer' Computing

Many applications fit a 'client-server' model

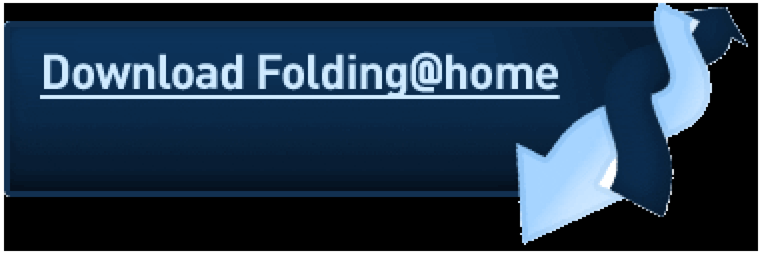
– 'it does not matter where the computer or data is' –

and if you have mainly compute tasks and little data,
even idle home PCs can contribute compute power
– although network bandwidth is limited …



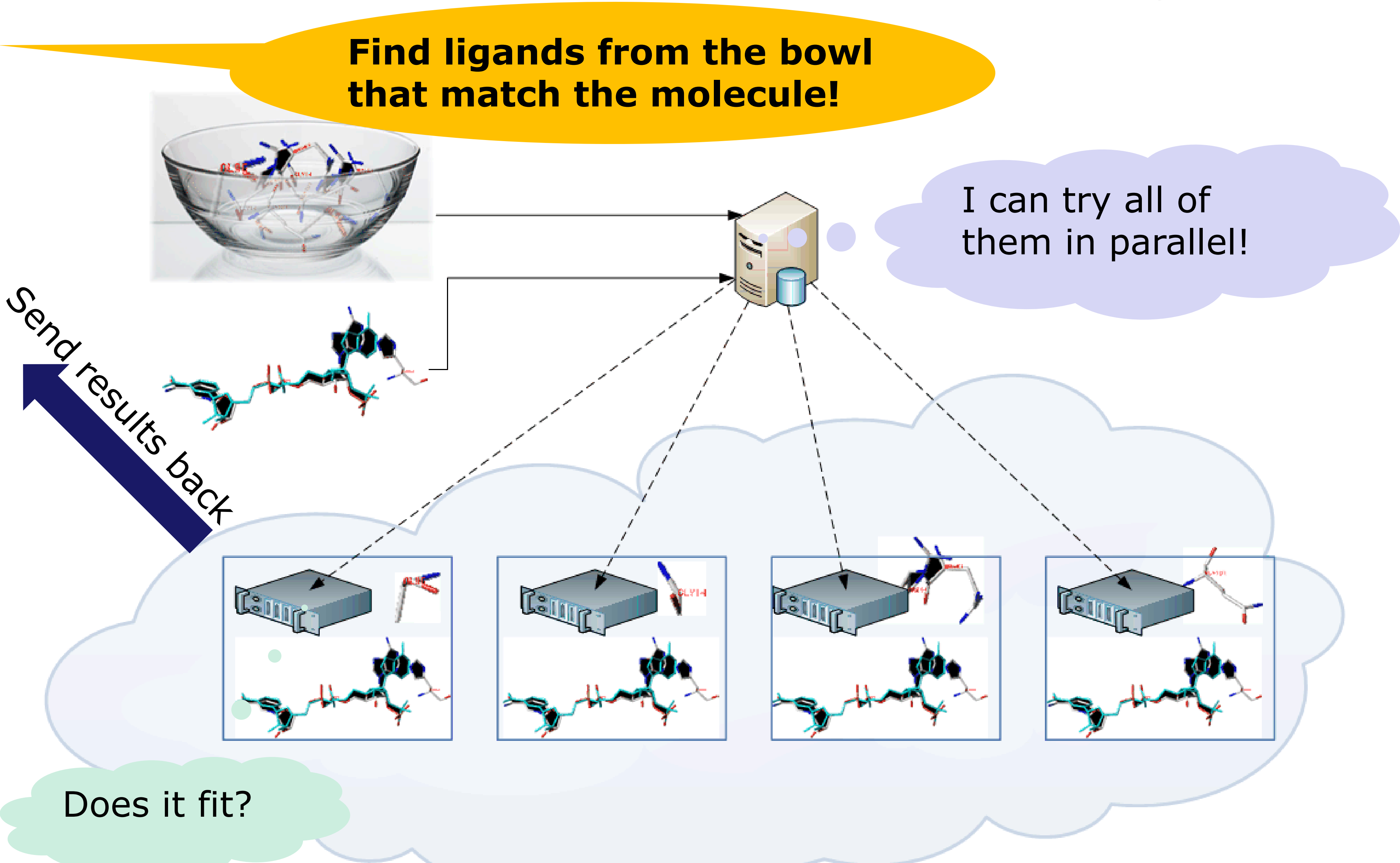
Pioneered ~ 1996 by SETI@home and 'distributed.net'

BOINC: generic middleware for 'volunteer' grids: 2005



go to boinc.berkeley.edu for information and links to projects

Conveniently Parallel Cluster Computing



Find ligands from the bowl that match the molecule!

I can try all of them in parallel!

Send results back

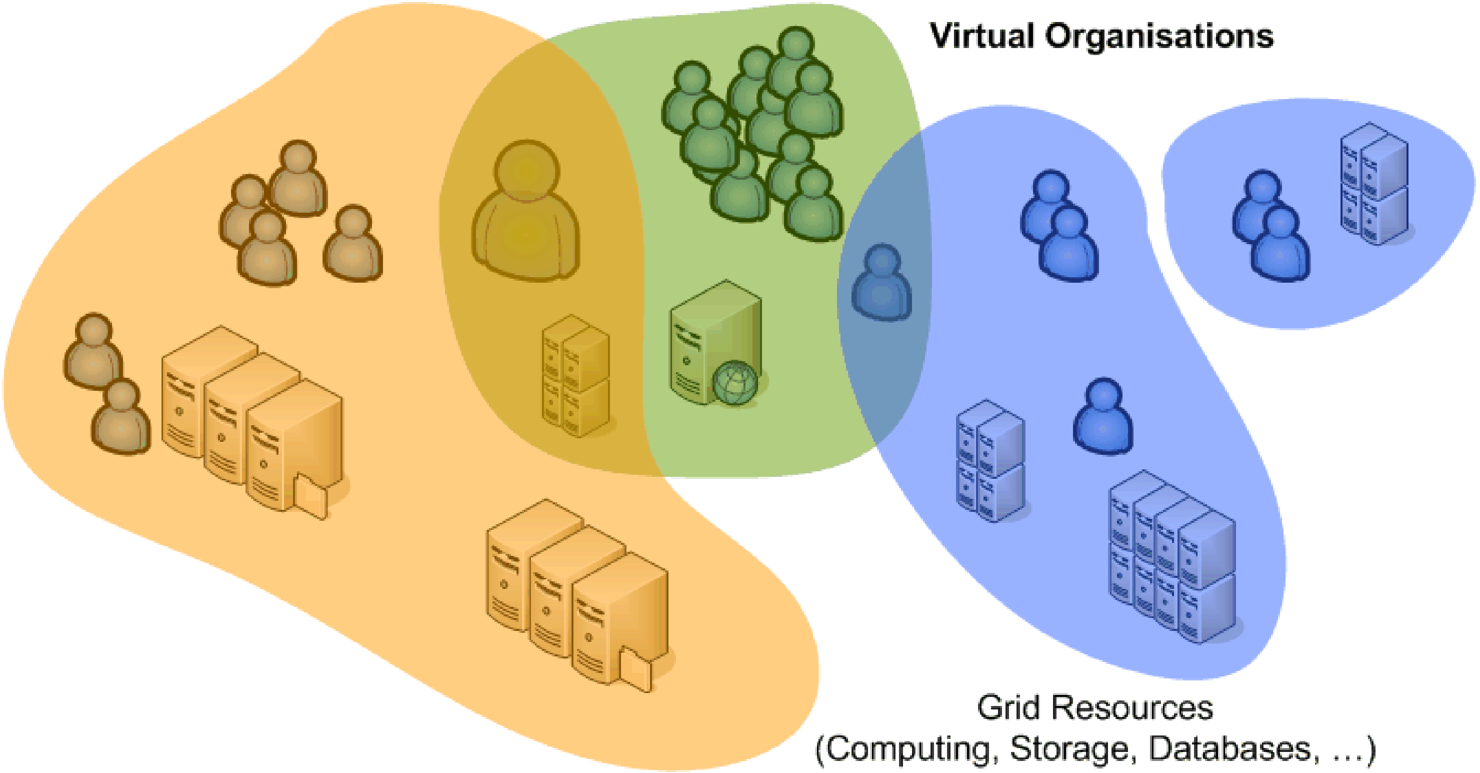
Does it fit?

Cross-domain and global grids

Today mostly science

The communities that make up the grid:

- **not under single hierarchical control**,
- temporarily **joining forces** to solve a particular problem at hand,
- bringing to the collaboration a subset of their resources,
- sharing those **at their discretion** and each **under their own conditions**.



Grid Infrastructure

To bring this about and sustain it requires a *persistent infrastructure* based on standards

Hardware infrastructure

clusters, supercomputers, databases, mass storage, visualisation, networks

Trust and AAA infrastructure

authentication, authorization, accounting, billing and settlement

Software infrastructure

execution services, workflow, resource information systems, database access, storage management, meta-data

Application infrastructure

user support, and ICT experts … with domain knowledge

BiG Grid
the dutch e·science grid

Nikhef (NDPF)

1200	processor cores
390 000	GByte disk
10 000	Mbps networks

SARA (GINA+LISA)

~3600	processor cores
950 000	GByte disk
2 000 000	GByte tape
4x 10 000	Mbps networks

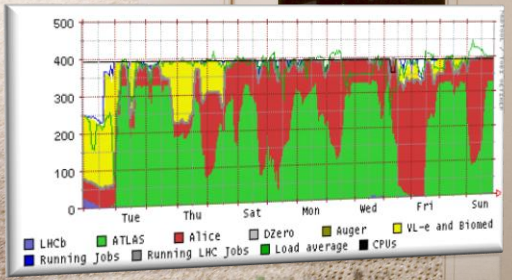
RUG-CIT (Grid)

~ 120	processor cores
8 800	GByte disk
10 000	Mbps networks

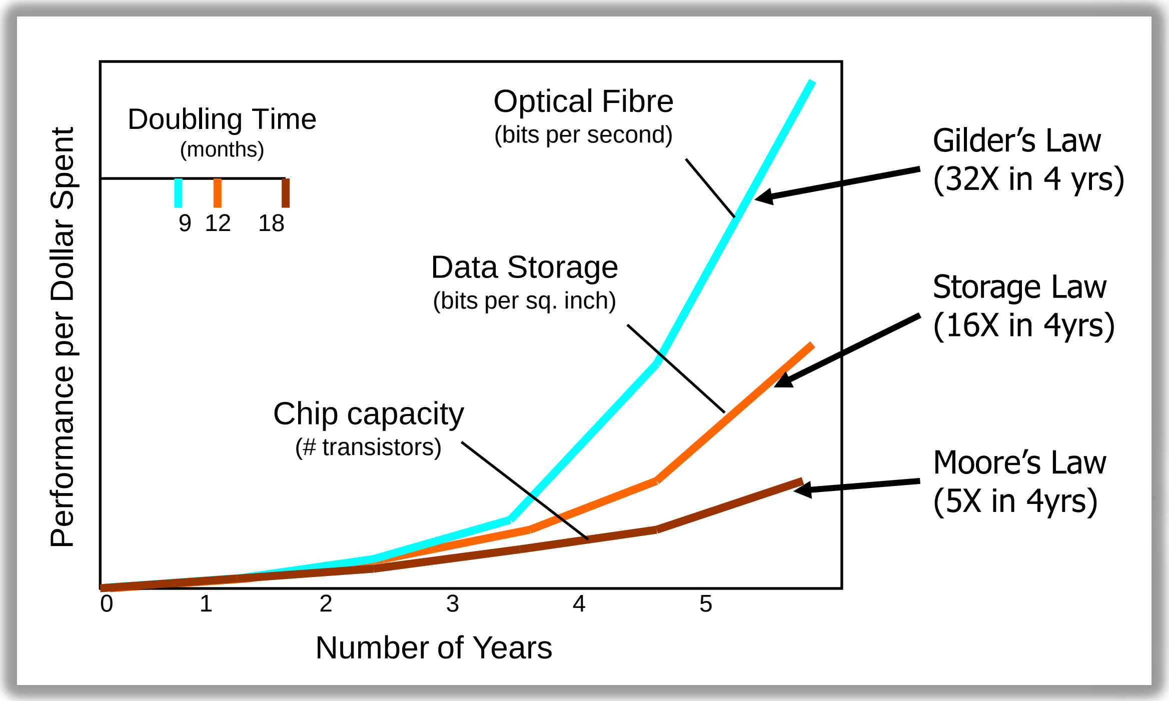
Philips Research Ehv

(planned 2008 Q2)

2000	processor cores
100 000	GByte disk
1 000	Mbps networks



There's always a network close to you



NL Light

SURF NET

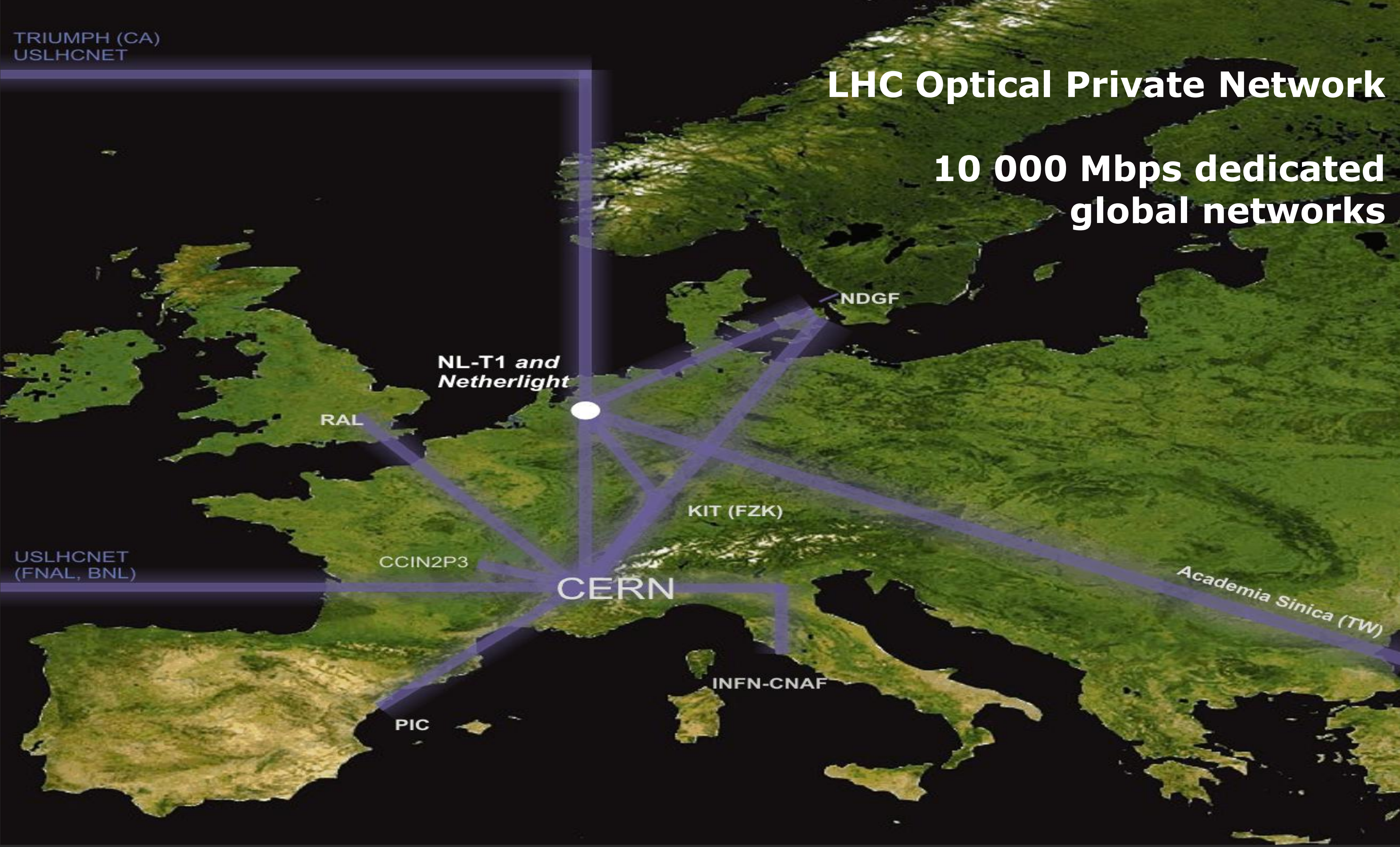
SURFnet pioneered 'lambda' and hybrid networks in the world

- and likely contributed to the creation of a market for 'dark fibre' in the Netherlands

There's always fibre within 2 miles from you – where ever you are!

(it's just that last mile to your home that's missing – and the business model of your telecom provider…)

Interconnecting the Grid – the Network



LHC Optical Private Network

10 000 Mbps dedicated global networks

Trust Infrastructure and Security

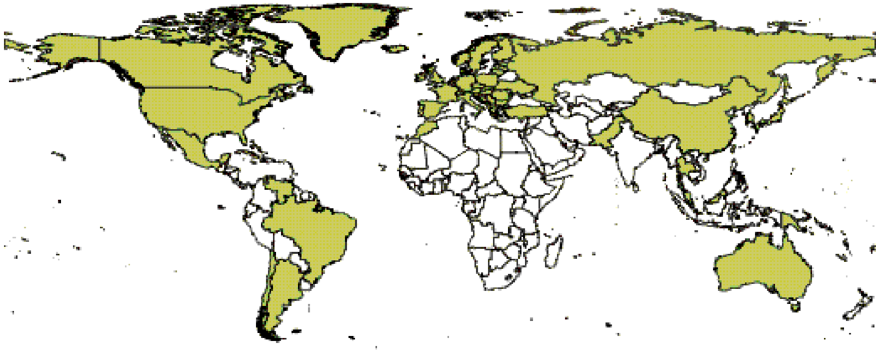
Why would I trust you? How do I know who you are?

'digital signatures and certificates be used as digital identities'

- even in Europe 1999/93/EC got only limited adoption
- In the Netherlands, 'Wet Digitale Handtekening 2003' for the general public was effectively superseded by DigiD – based on federation technology by SURFnet ...

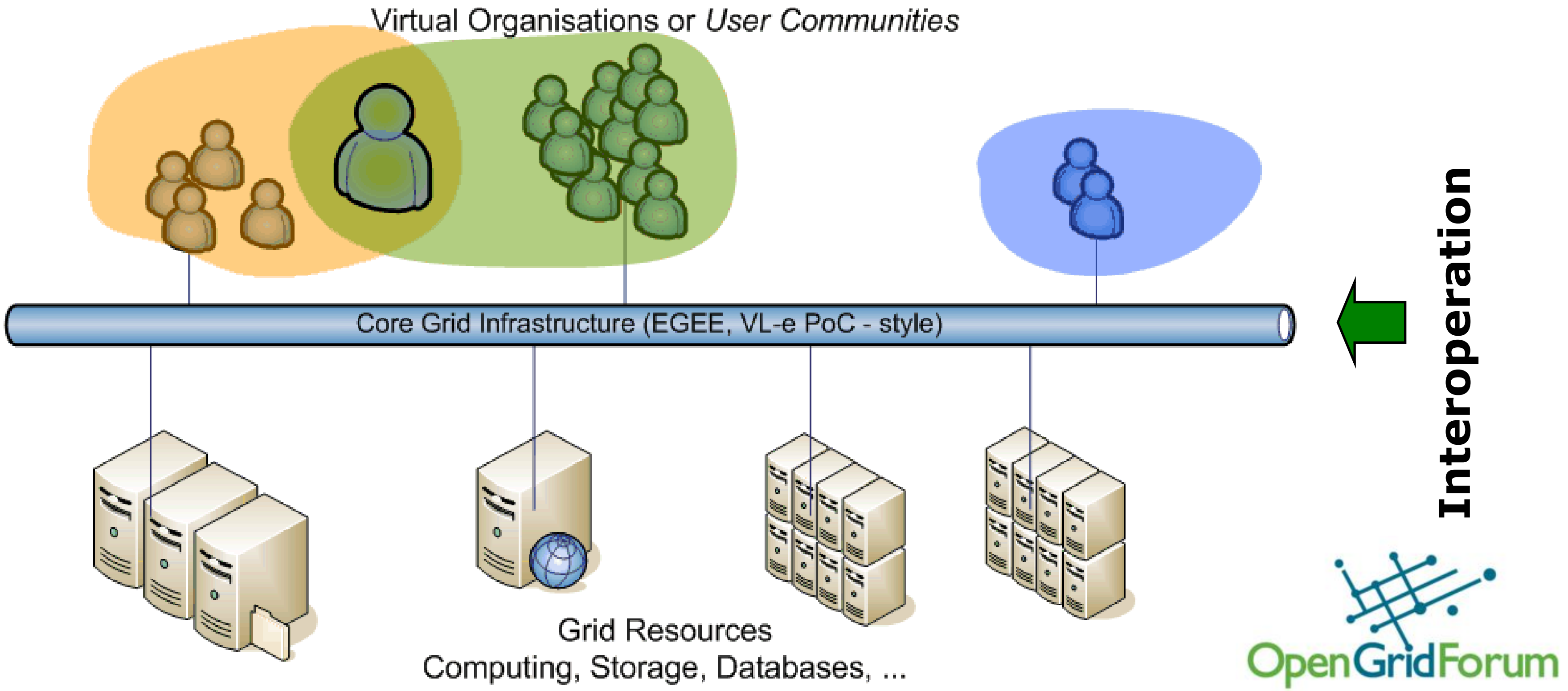
For the Grid a truly global identity is needed
–– so we built the International Grid Trust Federation

- supported by the EU and e-IRG delegates
- over 80 member Authorities



IGTF
International Grid Trust Federation
AP|EU|TAG

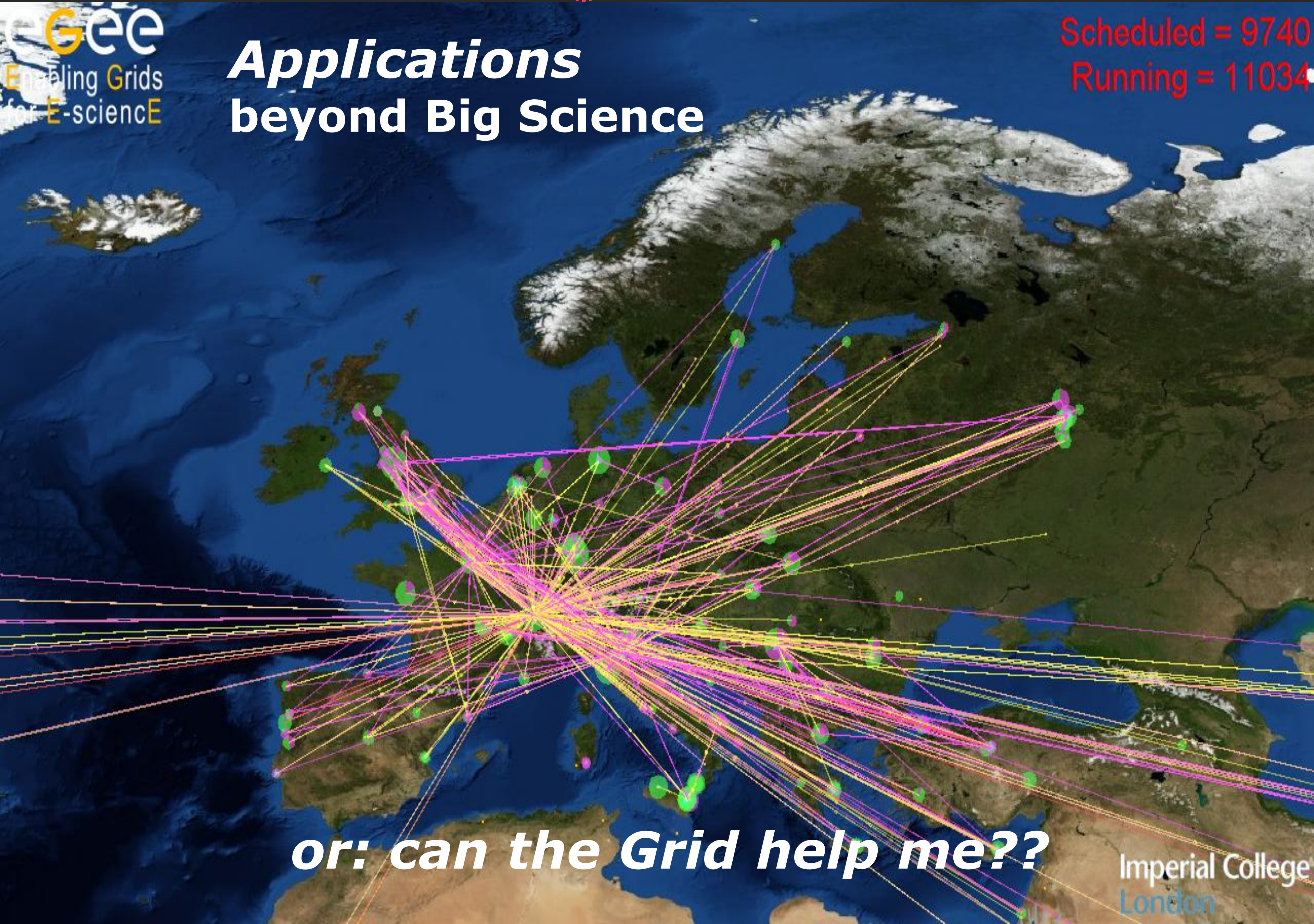
Software – connecting heterogeneous sources


Virtual Organisations or *User Communities*
Core Grid Infrastructure (EGEE, VL-e PoC - style)
Interoperation
Grid Resources
Computing, Storage, Databases, ...
OpenGridForum

- Use standards (like Web Services) to interoperate and prevent lock-in
- Use the experience of colleagues and best-of-breed solutions
- Connect to the infrastructure based on these open protocols

the web is a success because everyone agreed on 'http' and 'HTML'!

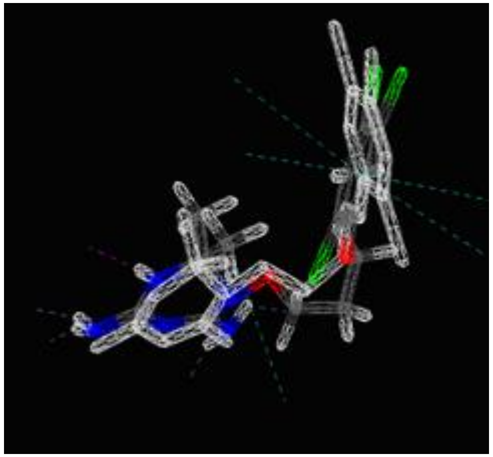
Applications
beyond Big Science



or: can the Grid help me??

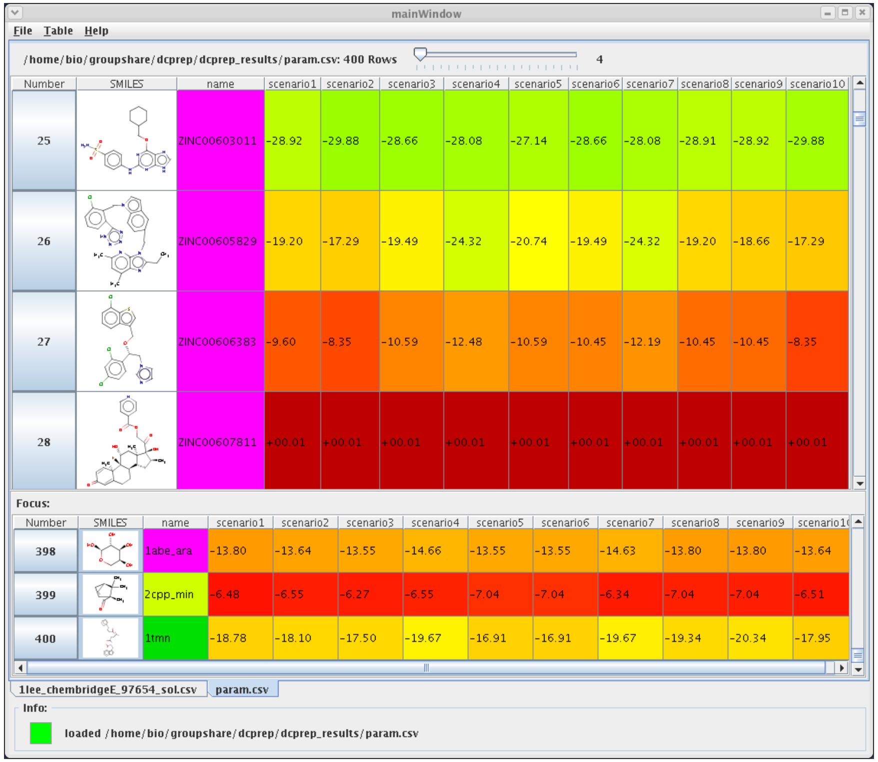
WISDOM: drug discovery

Wide-area In-Silico Docking On Malaria



over 46 million ligands virtually docked on malaria and H5N1 avian flu viruses in less than a month

100 *years* of work
on a single computer
sped-up about ~ 100 times!



/home/bio/groupshare/dcprep/dcprep_results/param.csv: 400 Rows 4

Number	SMILES	name	scenario1	scenario2	scenario3	scenario4	scenario5	scenario6	scenario7	scenario8	scenario9	scenario10
25		ZINC00603011	-28.92	-29.88	-28.66	-28.08	-27.14	-28.66	-28.08	-28.91	-28.92	-29.88
26		ZINC00605829	-19.20	-17.29	-19.49	-24.32	-20.74	-19.49	-24.32	-19.20	-18.66	-17.29
27		ZINC00606383	-9.60	-8.35	-10.59	-12.48	-10.59	-10.45	-12.19	-10.45	-10.45	-8.35
28		ZINC00607811	+00.01	+00.01	+00.01	+00.01	+00.01	+00.01	+00.01	+00.01	+00.01	+00.01

Focus:

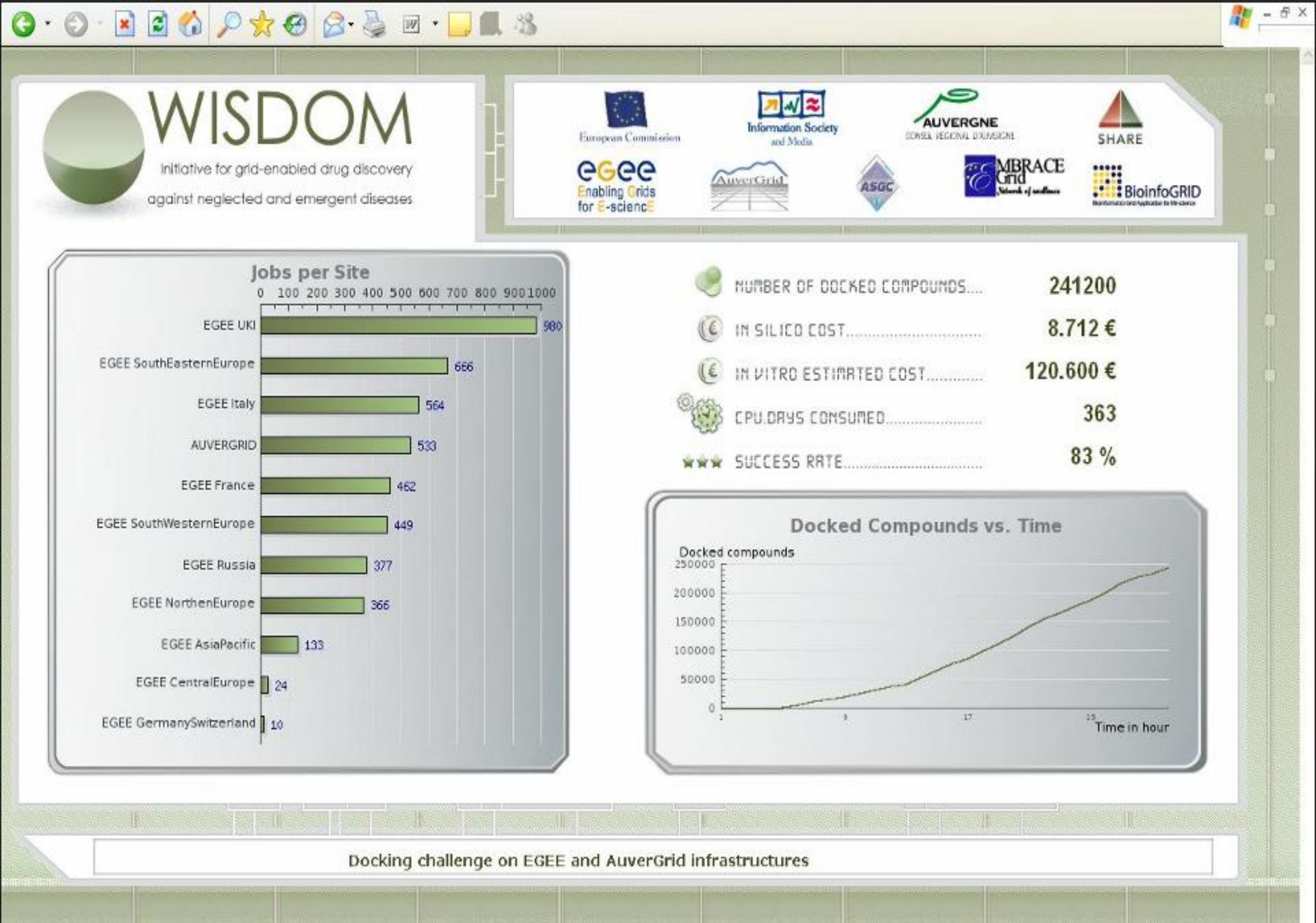
Number	SMILES	name	scenario1	scenario2	scenario3	scenario4	scenario5	scenario6	scenario7	scenario8	scenario9	scenario1(
398		1abe_ara	-13.80	-13.64	-13.55	-14.66	-13.55	-13.55	-14.63	-13.80	-13.80	-13.64
399		2cpp_min	-6.48	-6.55	-6.27	-6.55	-7.04	-7.04	-6.34	-7.04	-7.04	-6.51
400		1tmn	-18.78	-18.10	-17.50	-19.67	-16.91	-16.91	-19.67	-19.34	-20.34	-17.95

Info:

🟩 loaded /home/bio/groupshare/dcprep/dcprep_results/param.csv

- **47 sites**
- **15 countries**

- 3000 CPUs
- 12 TByte disk



WISDOM

Initiative for grid-enabled drug discovery against neglected and emergent diseases

NUMBER OF DOCKED COMPOUNDS....	241200
IN SILICO COST....................	8.712 €
IN VITRO ESTIMATED COST............	120.600 €
CPU.DAYS CONSUMED......................	363
SUCCESS RATE..........................	83 %

Docking challenge on EGEE and AuverGrid infrastructures

Science and Corporate Grids

big science is not alone

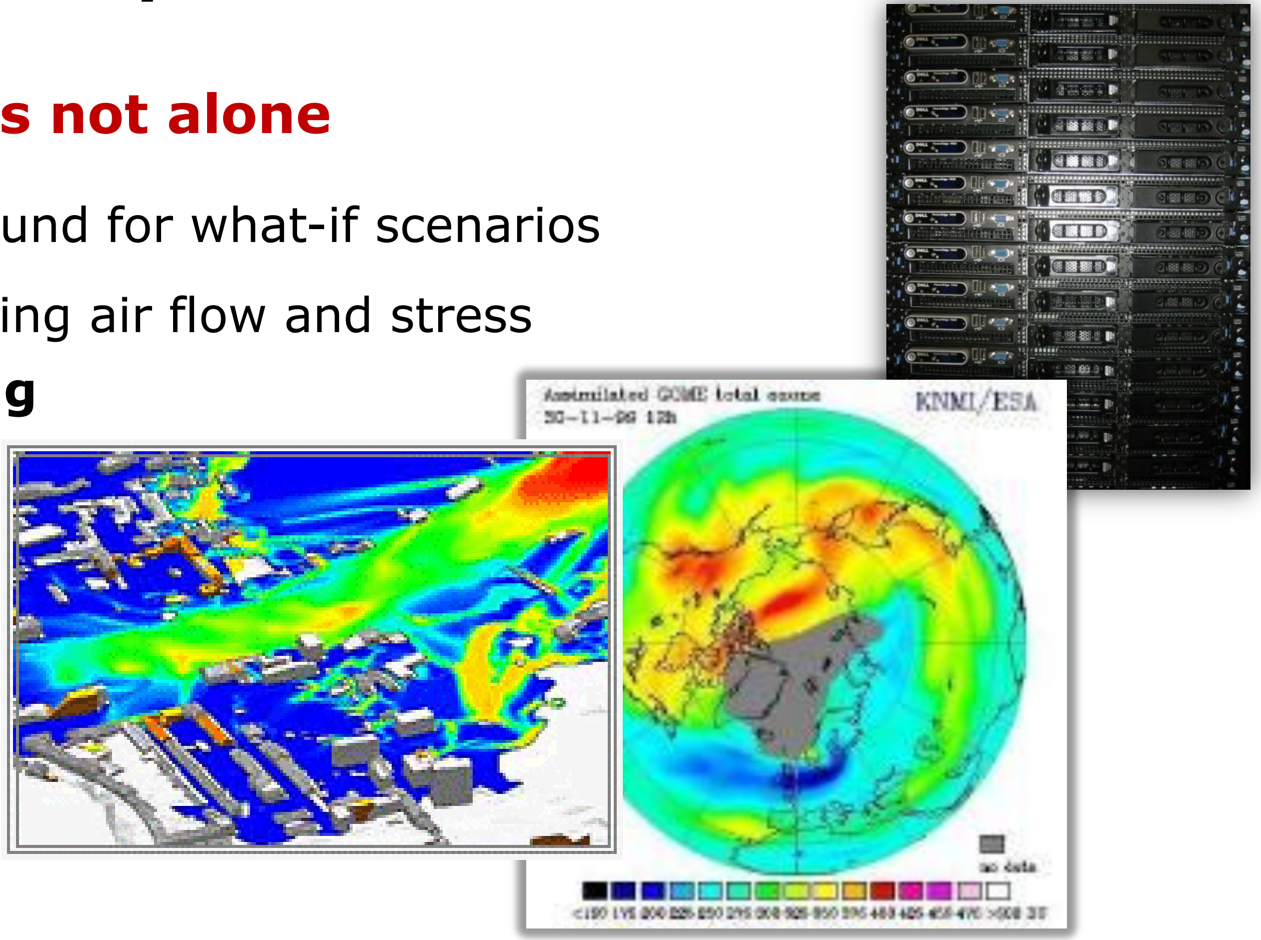
Finance rapid turn-around for what-if scenarios

Aerospace modelling air flow and stress

Medical imaging

Climate modelling

Flood prediction



But although the parallelism is convenient, managing complexity in a large-scale environment is not … and cooling and power constraints limit the data centre … the grid proposes the solution for advanced science

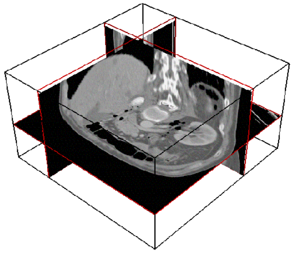
image sources: EGEE ESR VO, EGEE BioMedical VO, EU FP5 CrossGrid project, Nikhef

Image sources: VL-e Consortium Partners

Virtual Laboratory for e-Science

Data integration for genomics, proteomics, etc. analysis

Timo Breit et al.
Swammerdam Institute of Life Sciences

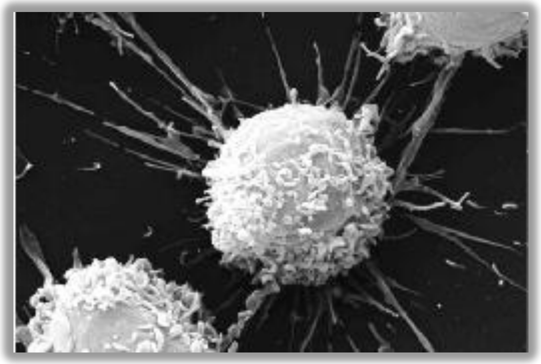


Medical Imaging and fMRI

Silvia Olabarriaga et al.
AMC and UvA IvI

Avian Alert and FlySafe

Willem Bouten et al.
UvA Institute for Biodiversity Ecosystem Dynamics, IBED



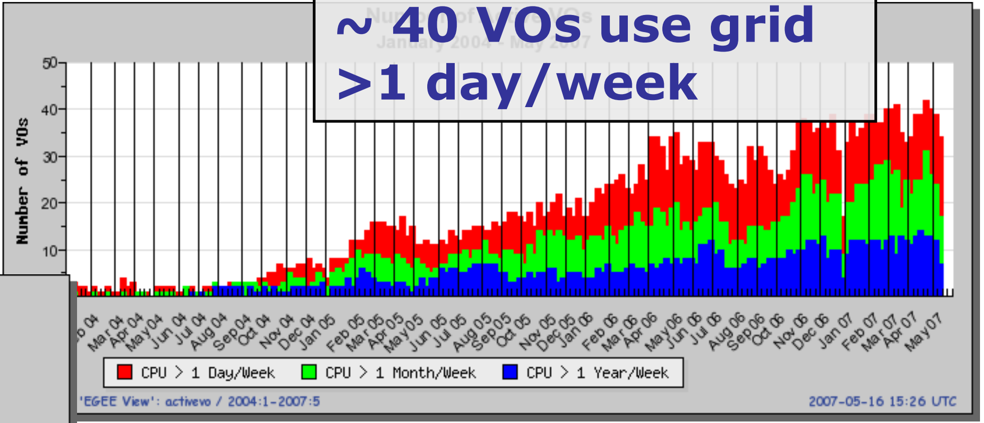
Bram Koster et al.
LUMC
Microscopic Imaging group

Molecular Cell Biology and 3D Electron Microscopy

Grid Infrastructures Work!

Enabling Grids for E-sciencE

Number of **active** VOs
in EU since 2004

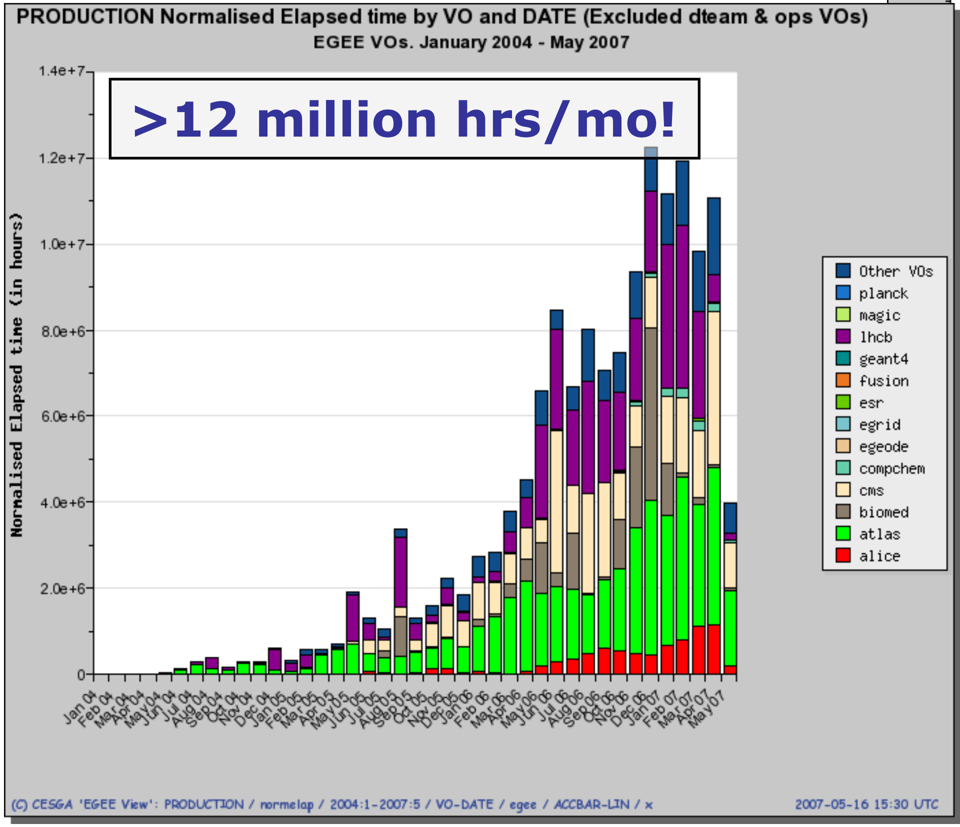


Number of active VOs
January 2004 - May 2007

'EGEE View': activevo / 2004:1-2007:5 2007-05-16 15:26 UTC

260 VOs total in EU
~ 40 VOs use grid
>1 day/week

Compute usage since 2004 by VO



PRODUCTION Normalised Elapsed time by VO and DATE (Excluded dteam & ops VOs)
EGEE VOs. January 2004 - May 2007

>12 million hrs/mo!

(C) CESGA 'EGEE View': PRODUCTION / normelap / 2004:1-2007:5 / VO-DATE / egee / ACCBAR-LIN / x 2007-05-16 15:30 UTC

BiG Grid
the dutch e-science grid

over 20 VOs hosted in NL

www.biggrid.nl

A reliable Grid Infrastructure needs operational support:

- availability monitoring
- reporting and follow-up
- user support

data: EGEE monitoring, RAL and CESGA, http://goc.grid-support.ac.uk/gridsite/accounting/

vl·e

BiG Grid
the dutch e·science grid

http://www.vl-e.nl/
http://www.biggrid.nl/
http://www.nikhef.nl/grid/

NIKHEF pdp