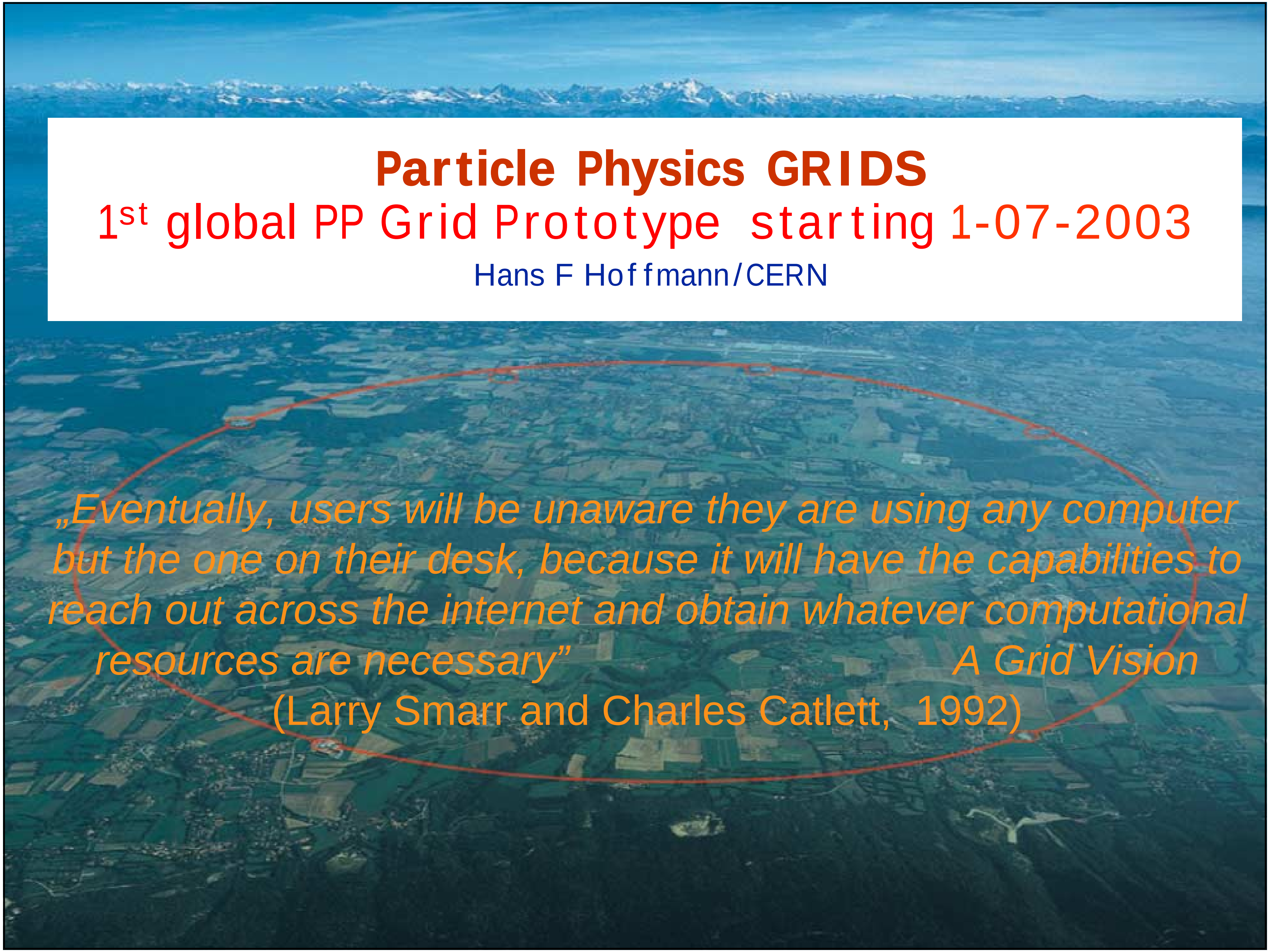
Particle Physics GRIDS

1st global PP Grid Prototype starting 1-07-2003

Hans F Hoffmann/CERN

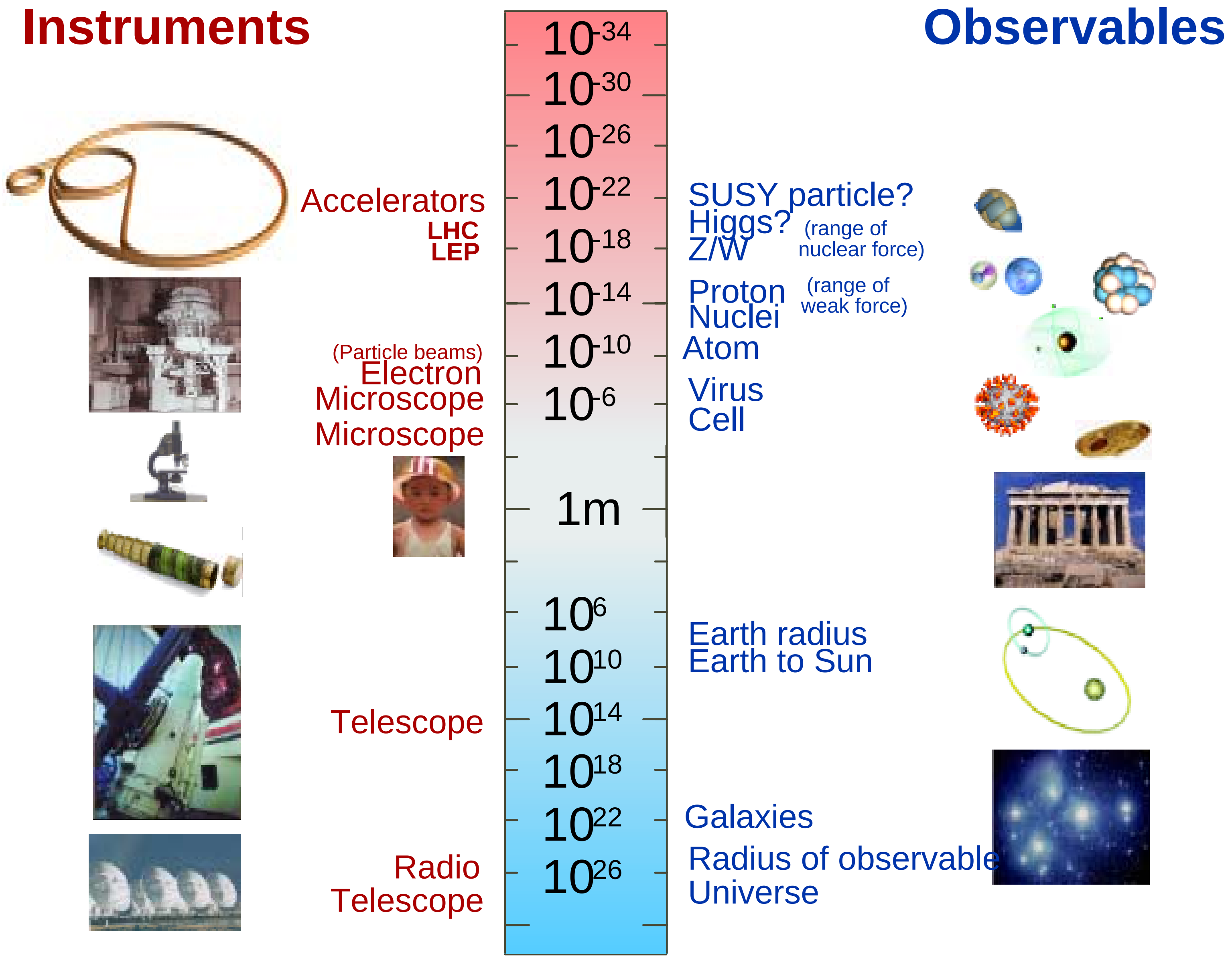


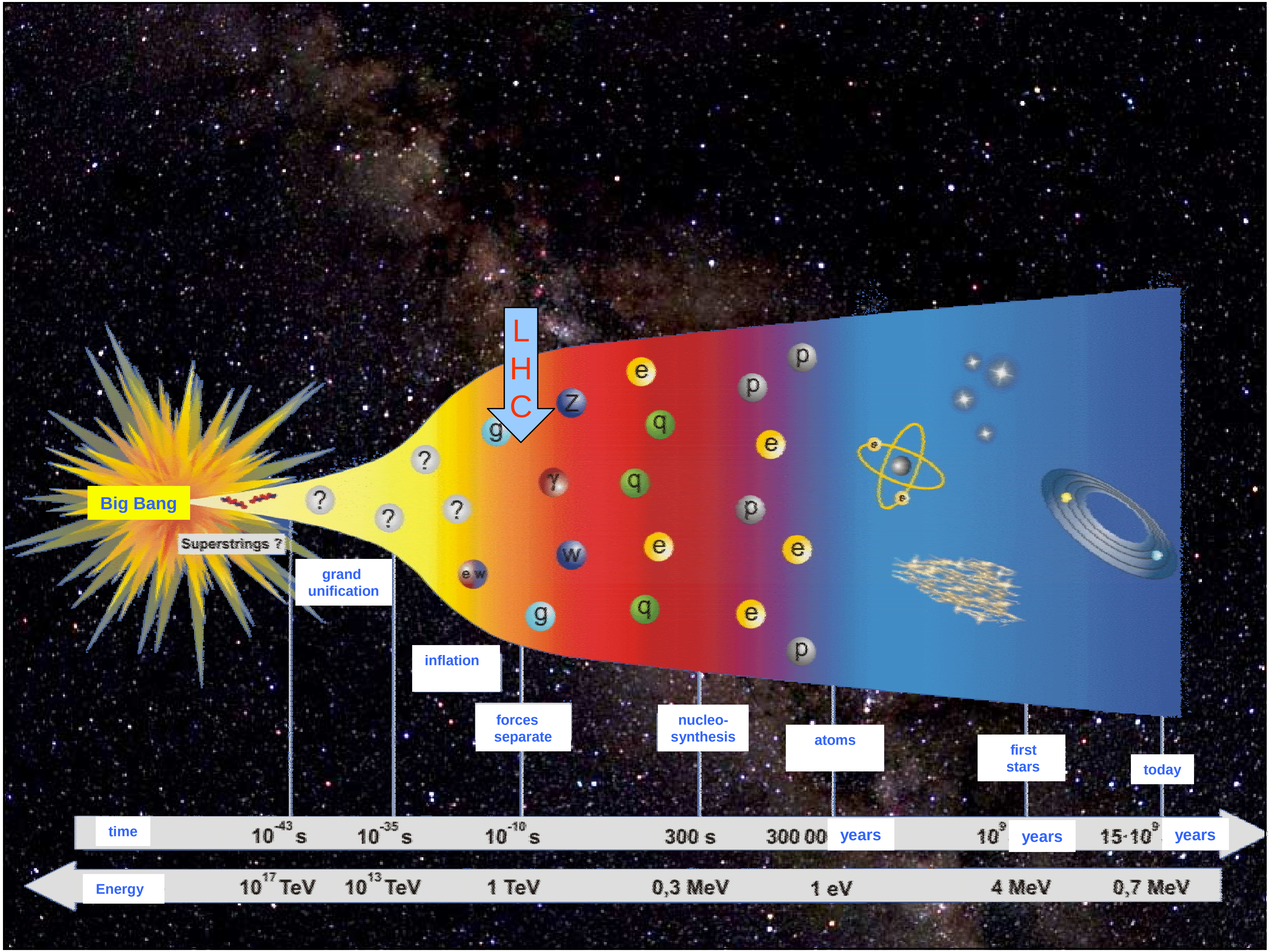
„Eventually, users will be unaware they are using any computer but the one on their desk, because it will have the capabilities to reach out across the internet and obtain whatever computational resources are necessary" *A Grid Vision*

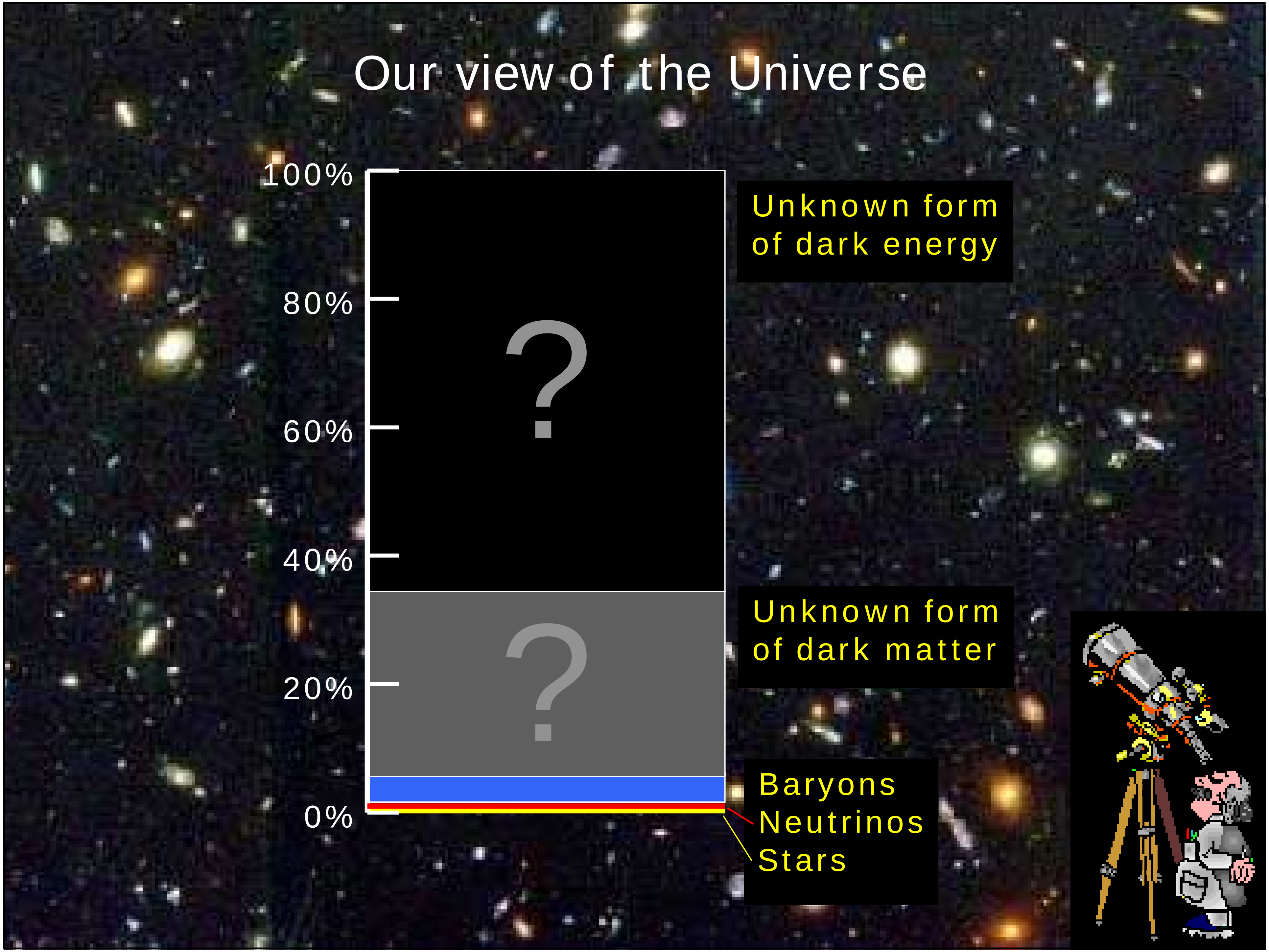
(Larry Smarr and Charles Catlett, 1992)

The size of the things

$\lambda = 1/p$




Big Bang
Superstrings ?
L
H
C
grand
unification
inflation
forces
separate
nucleo-
synthesis
atoms
first
stars
today
time
10^{-43} s
10^{-35} s
10^{-10} s
300 s
300 00 years
10^{9} years
$15 \cdot 10^{9}$ years
Energy
10^{17} TeV
10^{13} TeV
1 TeV
0,3 MeV
1 eV
4 MeV
0,7 MeV

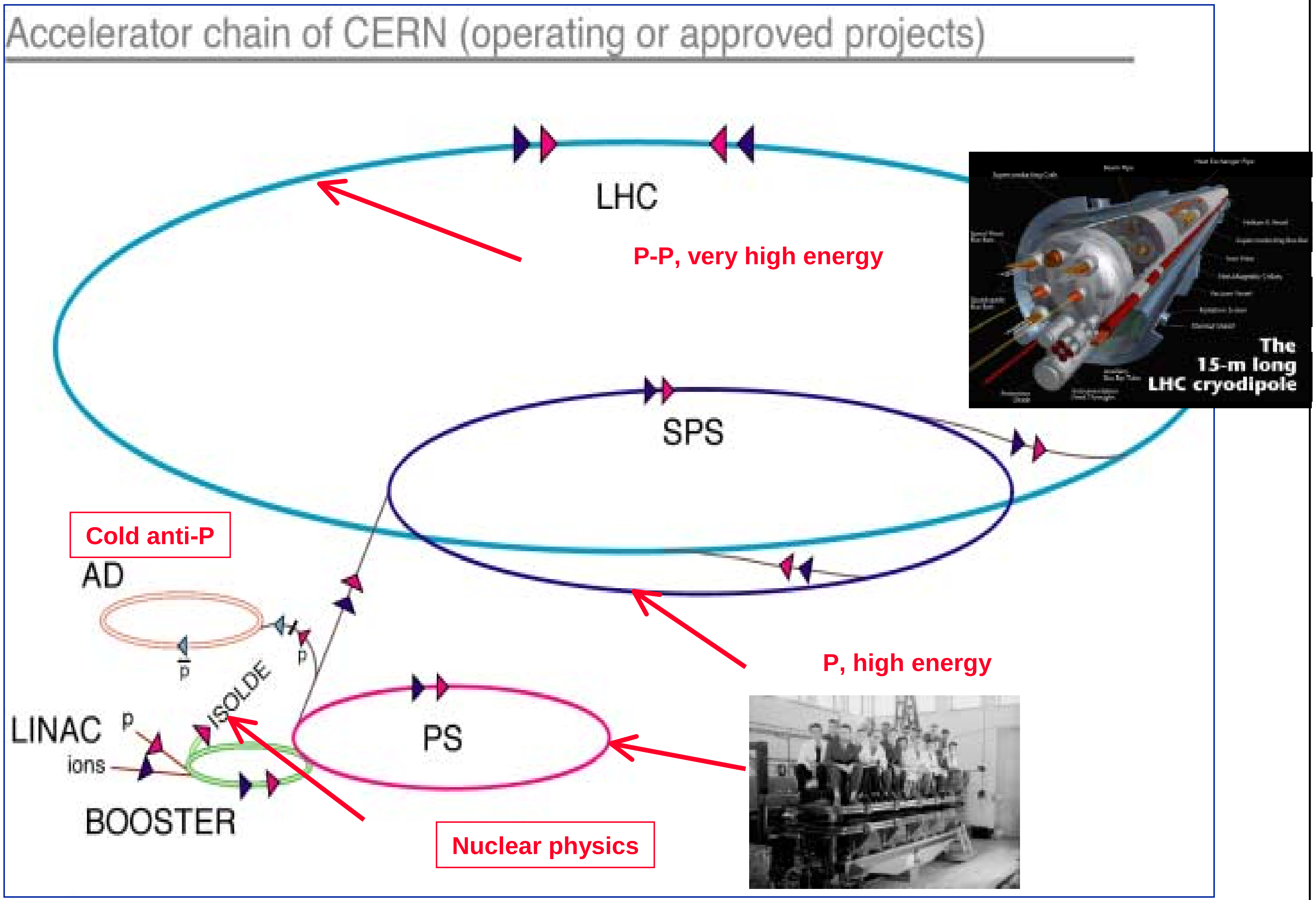


Our view of the Universe
100%
80%
60%
40%
20%
0%
?
?
Unknown form
of dark energy
Unknown form
of dark matter
Baryons
Neutrinos
Stars

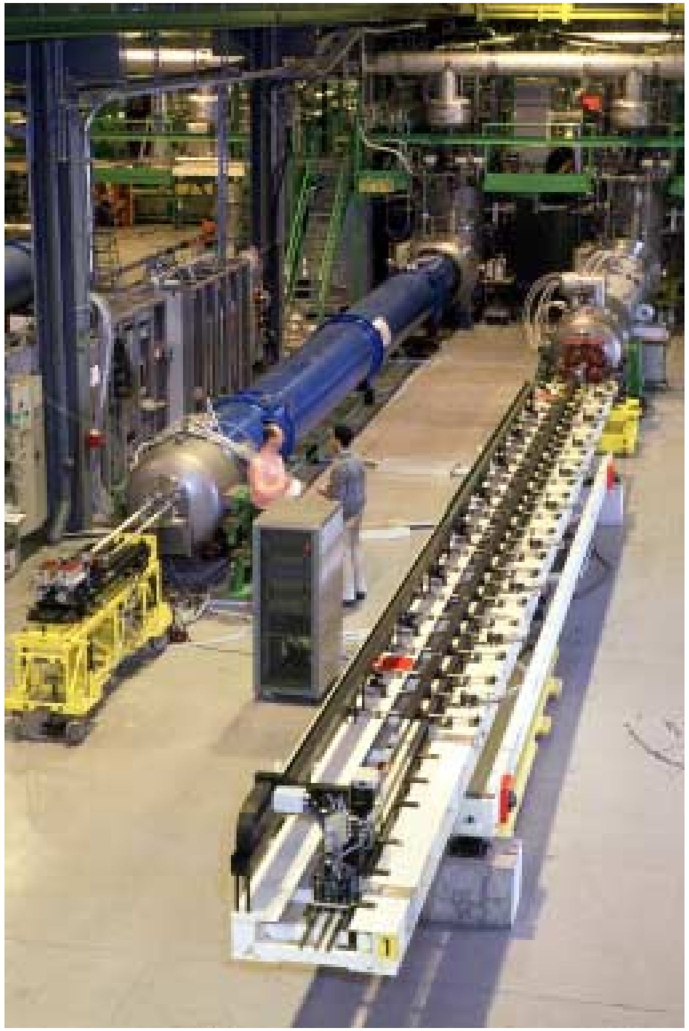
CERN Member States

~600 M€ yearly budget; ~2300 staff; >6000 scientists from 60 countries

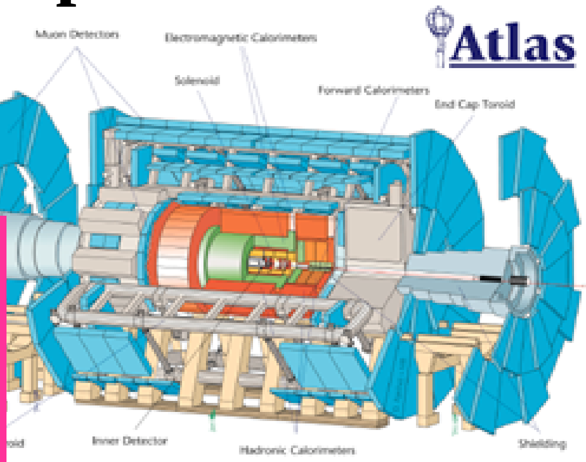
Scientific Instruments: Accelerators

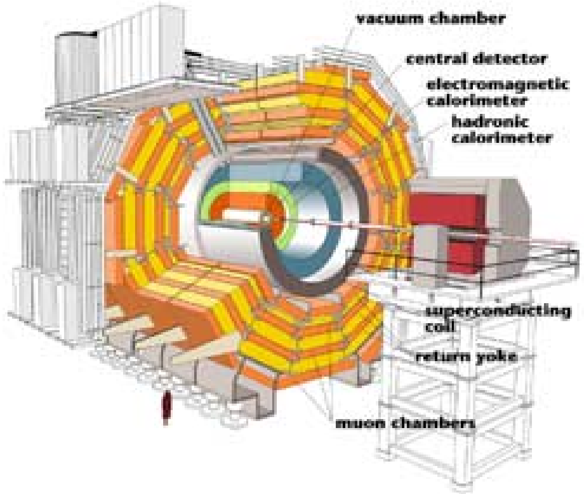


Superconducting magnets for the LHC on the test bench: dipole (CERN) and quadrupoles (CEA+IN2P3)



LHC Experiments





ATLAS, CMS:

- Higgs boson(s)

- SUSY particles

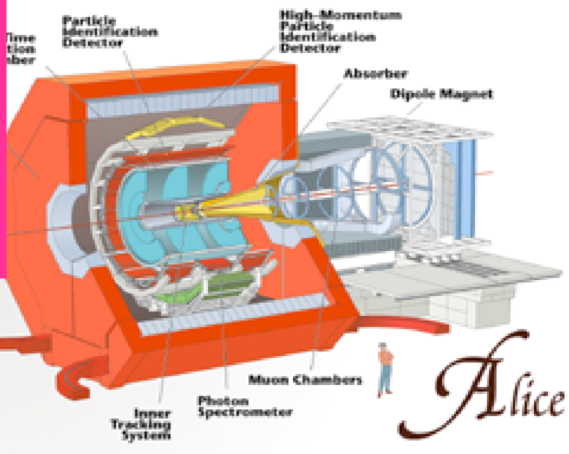
- …??

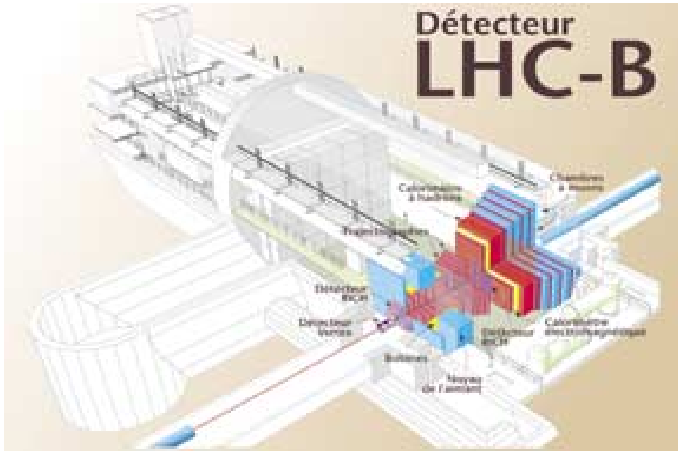
ALICE:

Quark Gluon Plasma

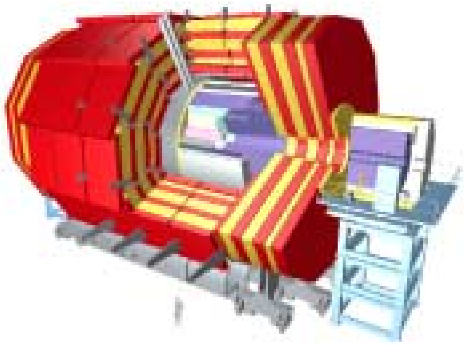
LHC-B:

- CP violation in B



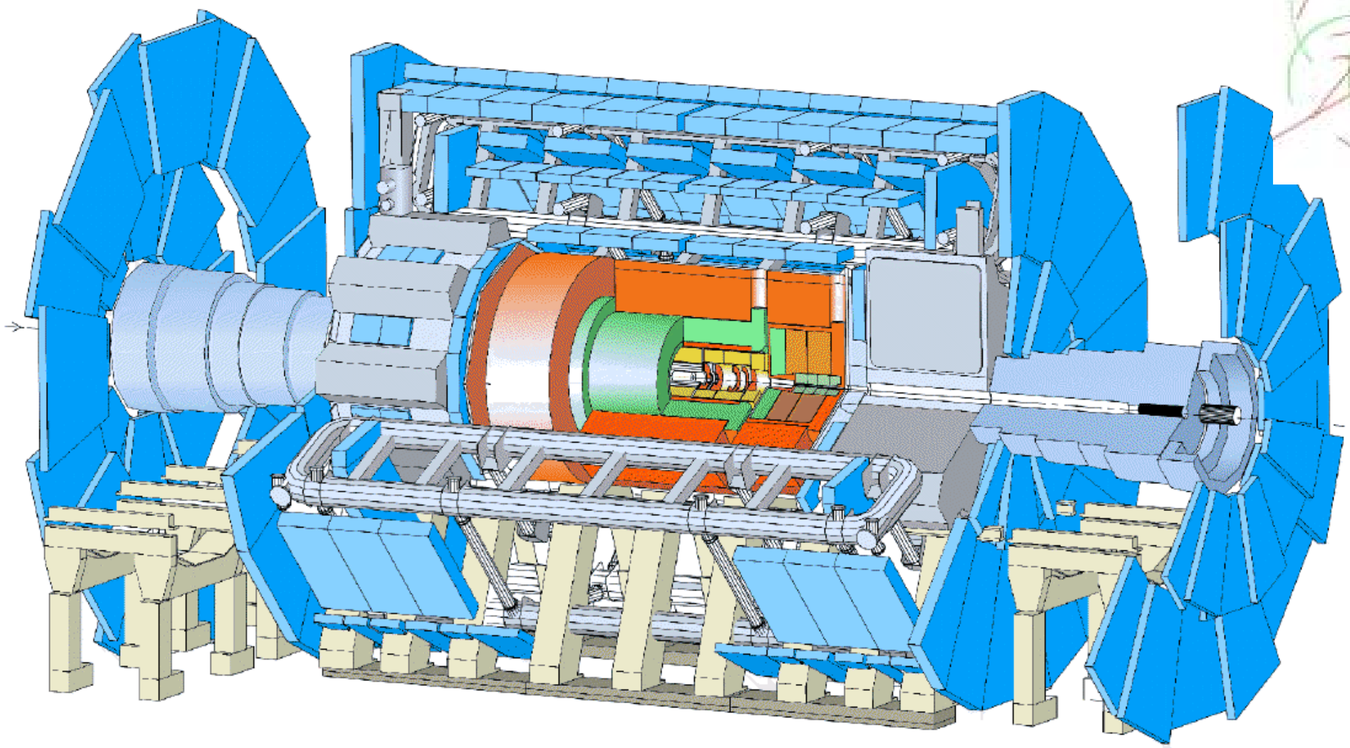


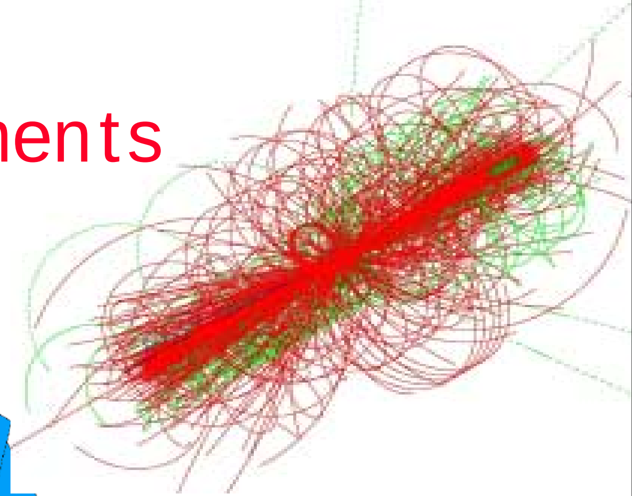
CMS :MagneticYoke on CERN site



ECAL: 40 Submodules assembled

Scientific Instruments: Experiments





10^9 events/s; 1 Petabyte/s at front-end electronics; 100MB/sec data rate;

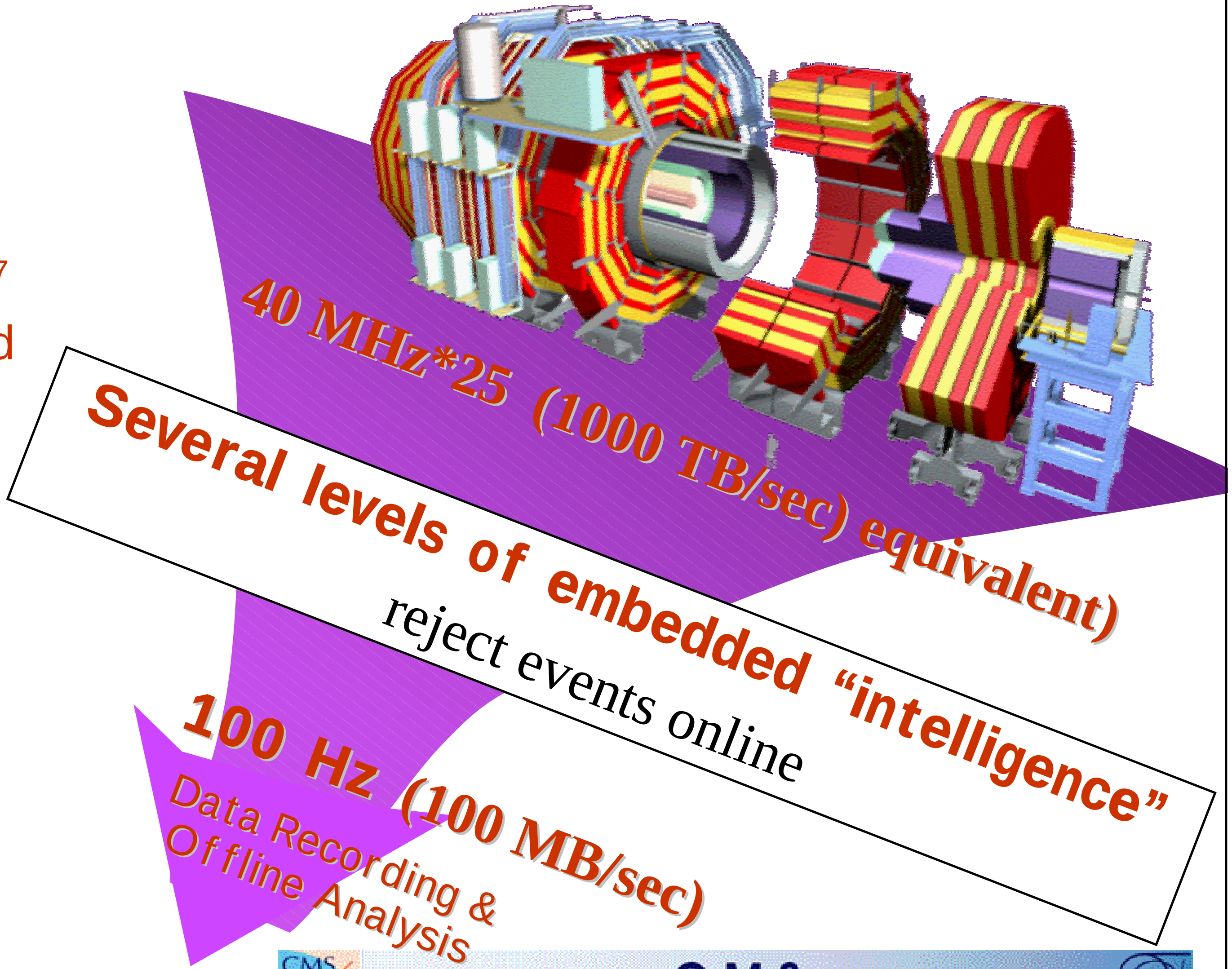
150 million sensors, 10 million lines of code; selectivity $1/10^{12}$; . . .

2000 collaborators from 150 institutes around the world

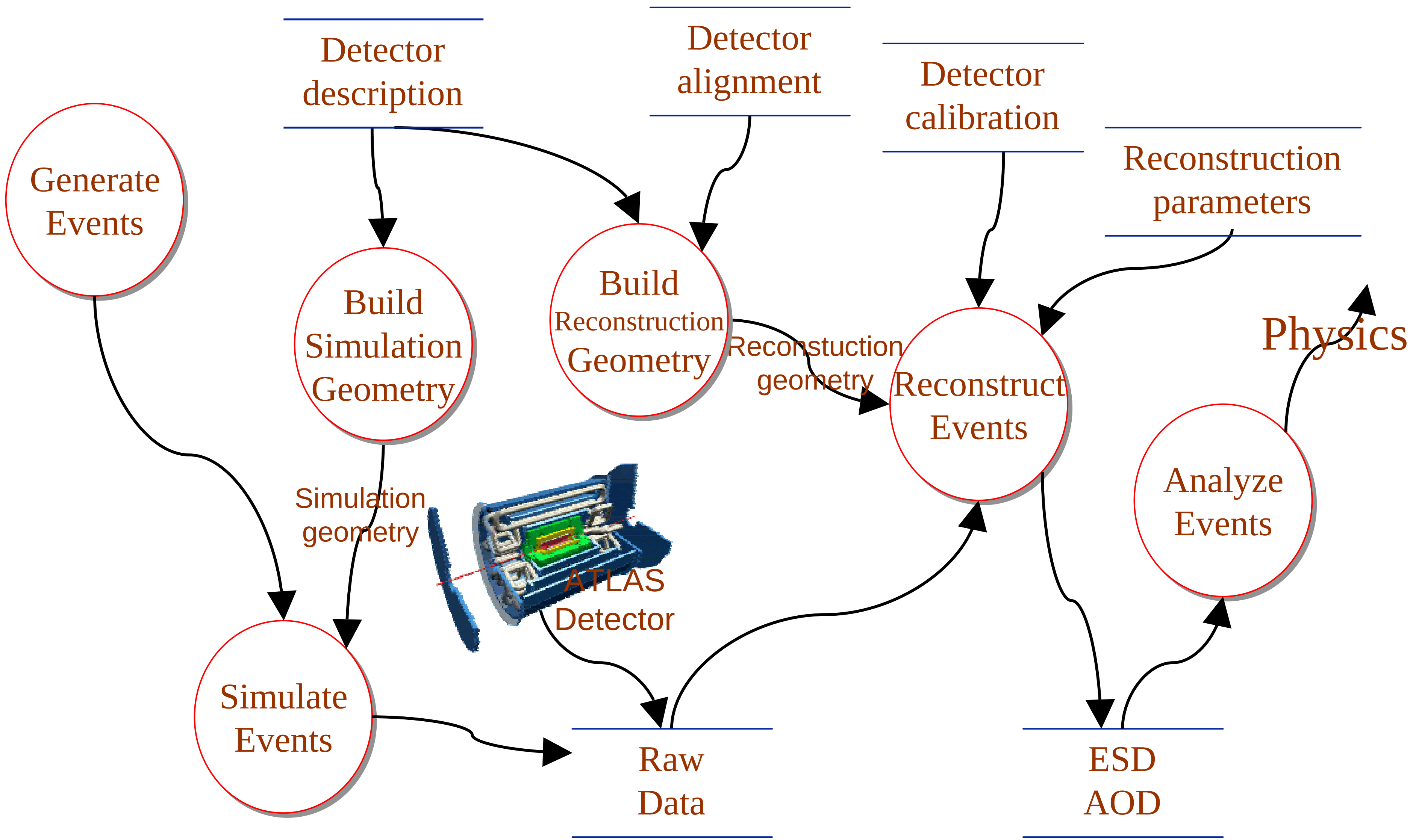
Tools: Data Acquisition & Triggers

Vast amounts of data

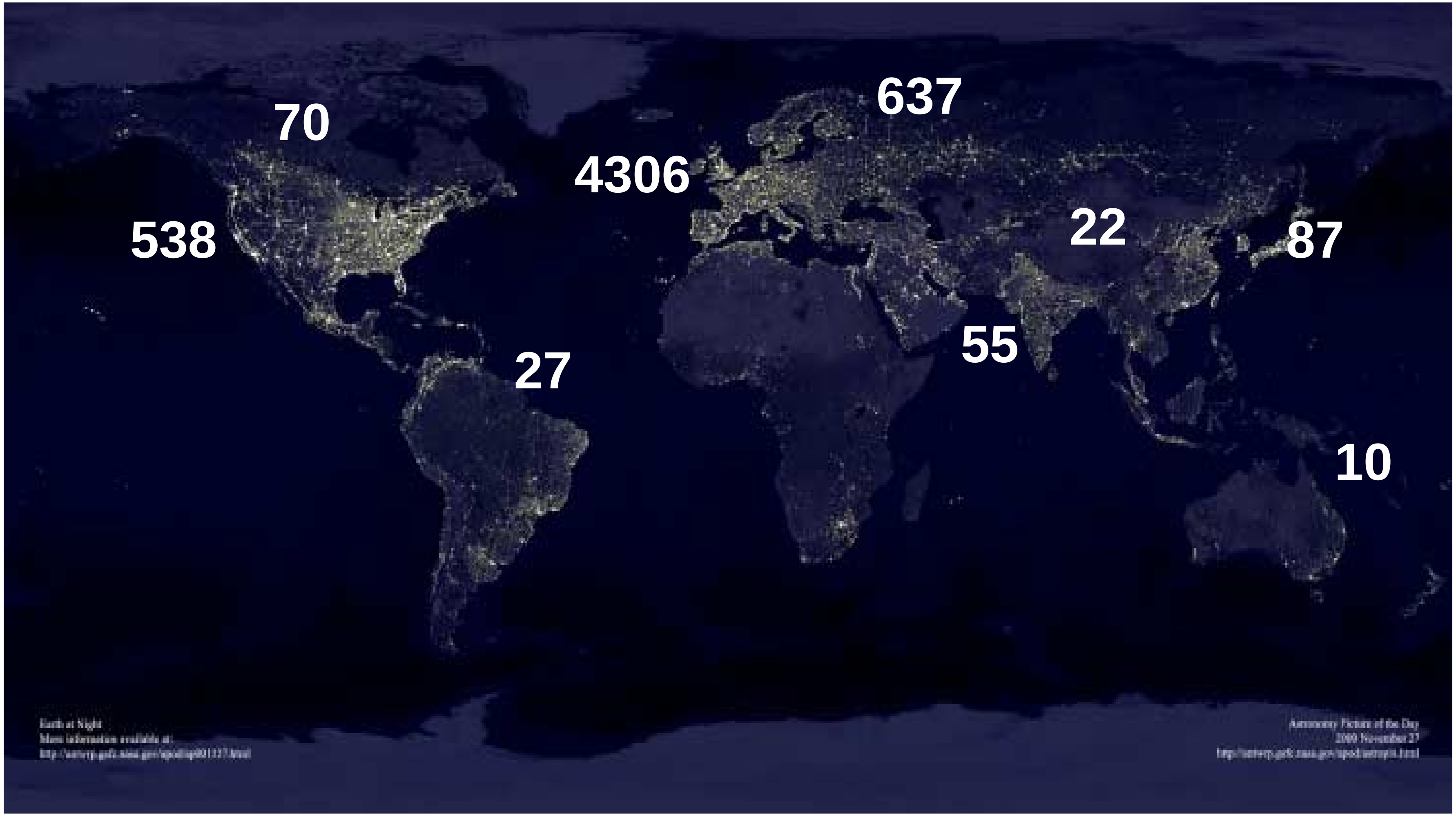
- Online reduction 10^7
- Keep highly selected interesting events



Offline Software, Data Curation



CERN's Users and Collaborating Institutes

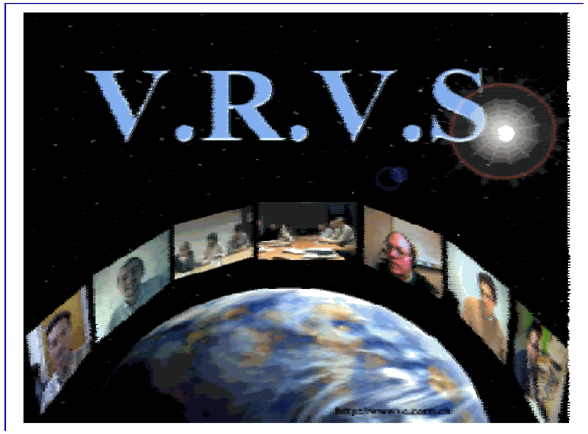


CERN's Network in the World

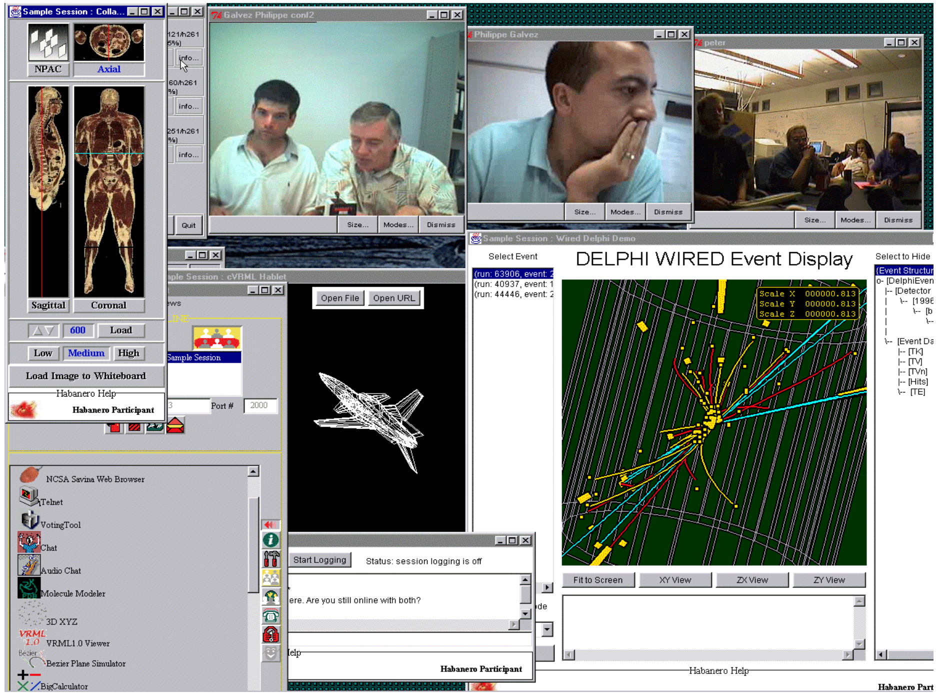
Strongly based in universities, 270 institutes in Europe, 4600 scientists, 210 institutes elsewhere, 1650 scientists; many students, postdocs; Annual throughput of ~300 engineers and ~800 physicists, 1 LHC experiment: ~1800 scientists

“Virtual Room” Videoconferencing System

HEP community
“virtual laboratories, - organisations”



6100 Hosts;
7 Internet2 Reflectors
Users In 52 Countries
Annual Growth 250%
http://vrvs.cern.ch/



LHC Collaborations - virtual, global Organisations

Clear common mission (~30 y for LHC), clear objectives

Open, world-wide, critical mass in all scientific and technological fields required, able to deal with all problems posed, free exchange of ideas, technologies, R&D, deal with IPR when required

"Lean, bottom-up“, democratic, self-organisation; success based on common goals and competition, recognition of individual contributions

Good record of achievements in terms of delivery to specs, schedules, budgets

CERN LHC programme re-baseline in 2001: 20% cost, 18 months schedule overruns since 1994, total cost to CERN ~3 BEuro

Peer Reviews: technical, scientific, organisational issues, . . .

Resources Reviews: regular reporting/reviews with all funding agencies involved, based on MoUs, meaning “best efforts” to fulfill objectives

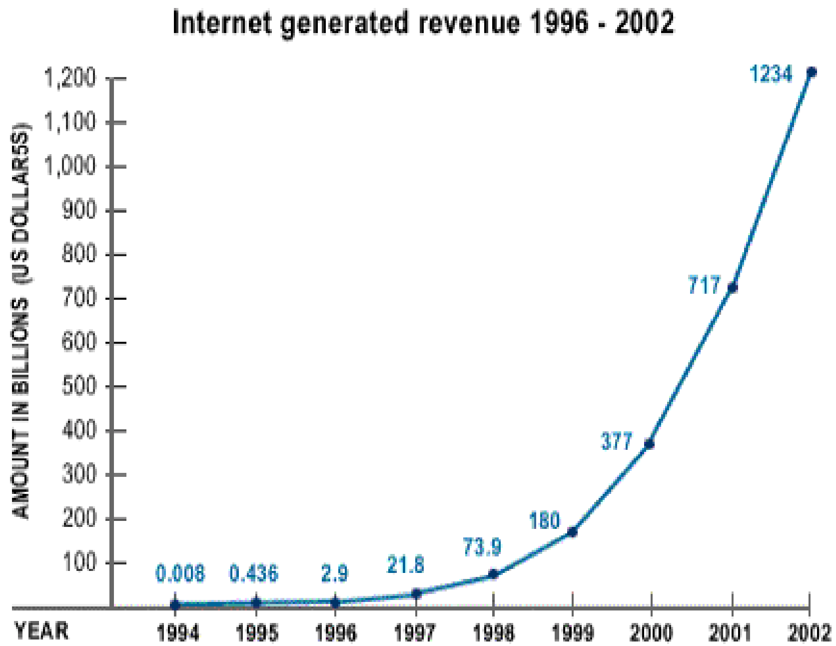
By-product of Particle Physics@CERN:
WWW and "infocern", the 1st web address ~1990
html (xml) open standards

D O Williams

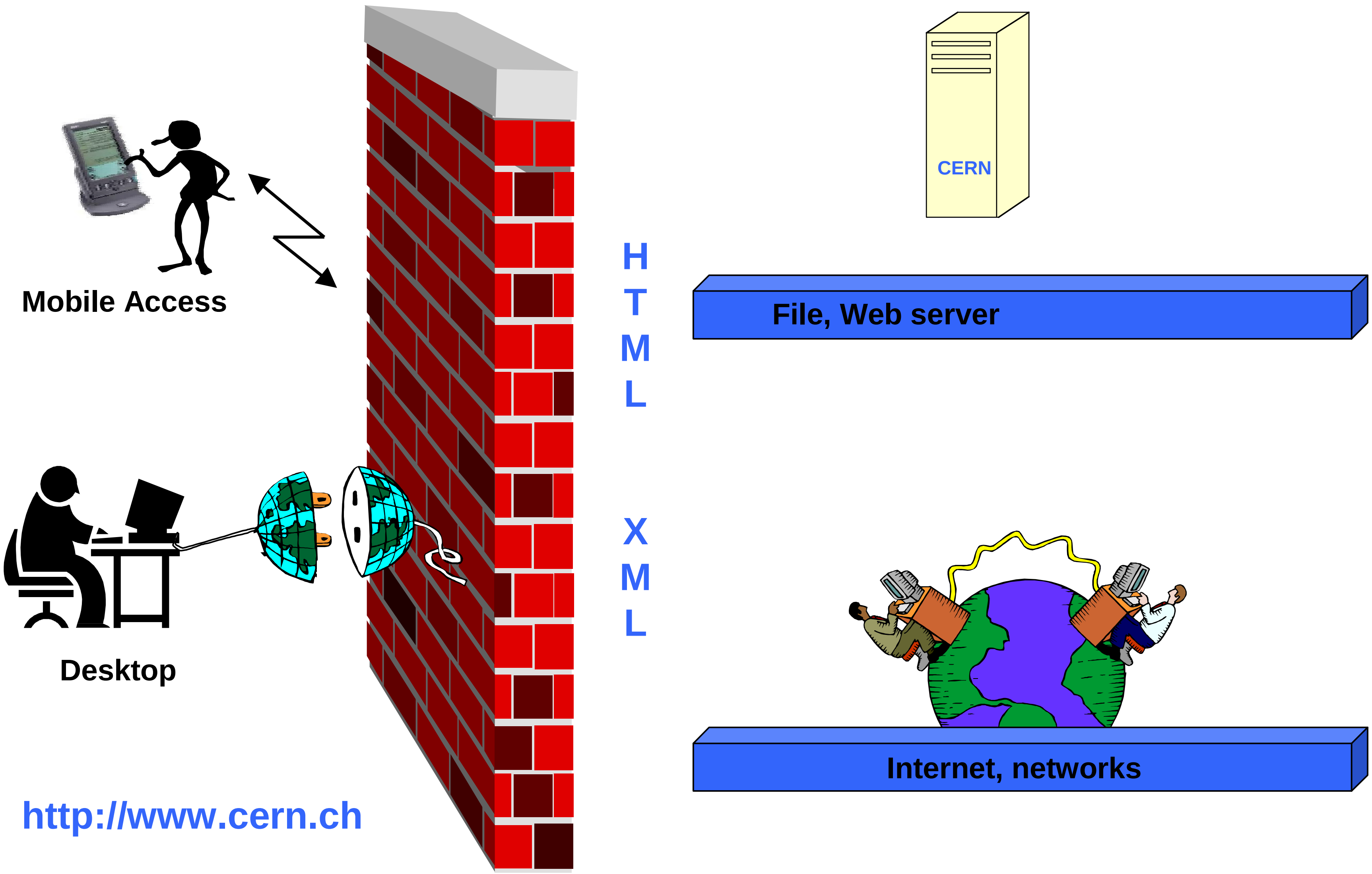
A great achievement and a fantastic idea, at the right time, making the internet available to everybody

It proves something about the benefits of assembling together urgent needs, infrastructure and smart people, and letting them interact..

And why we should not always listen to wise people who tell us that industry will always do better than we will....

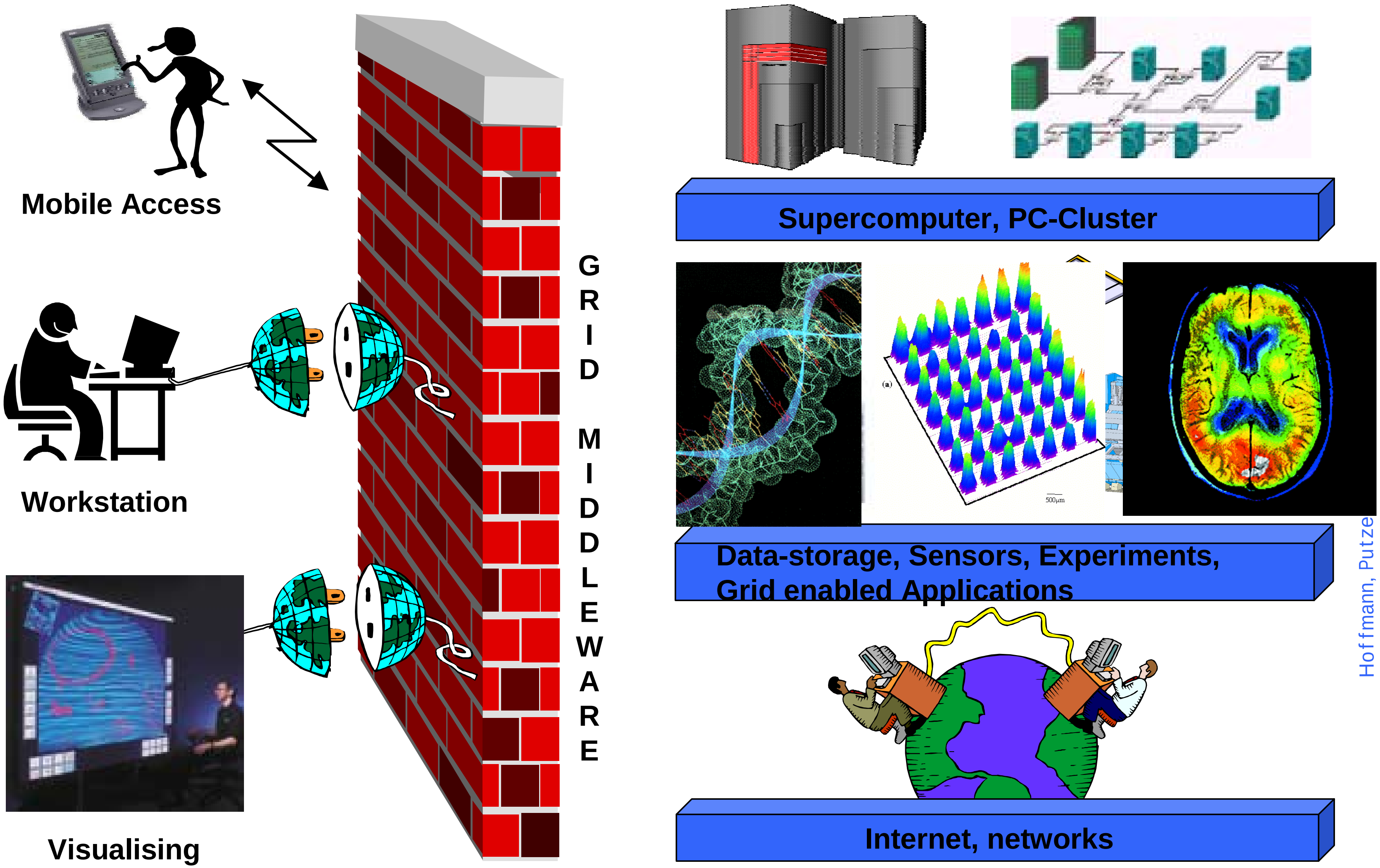


WWW

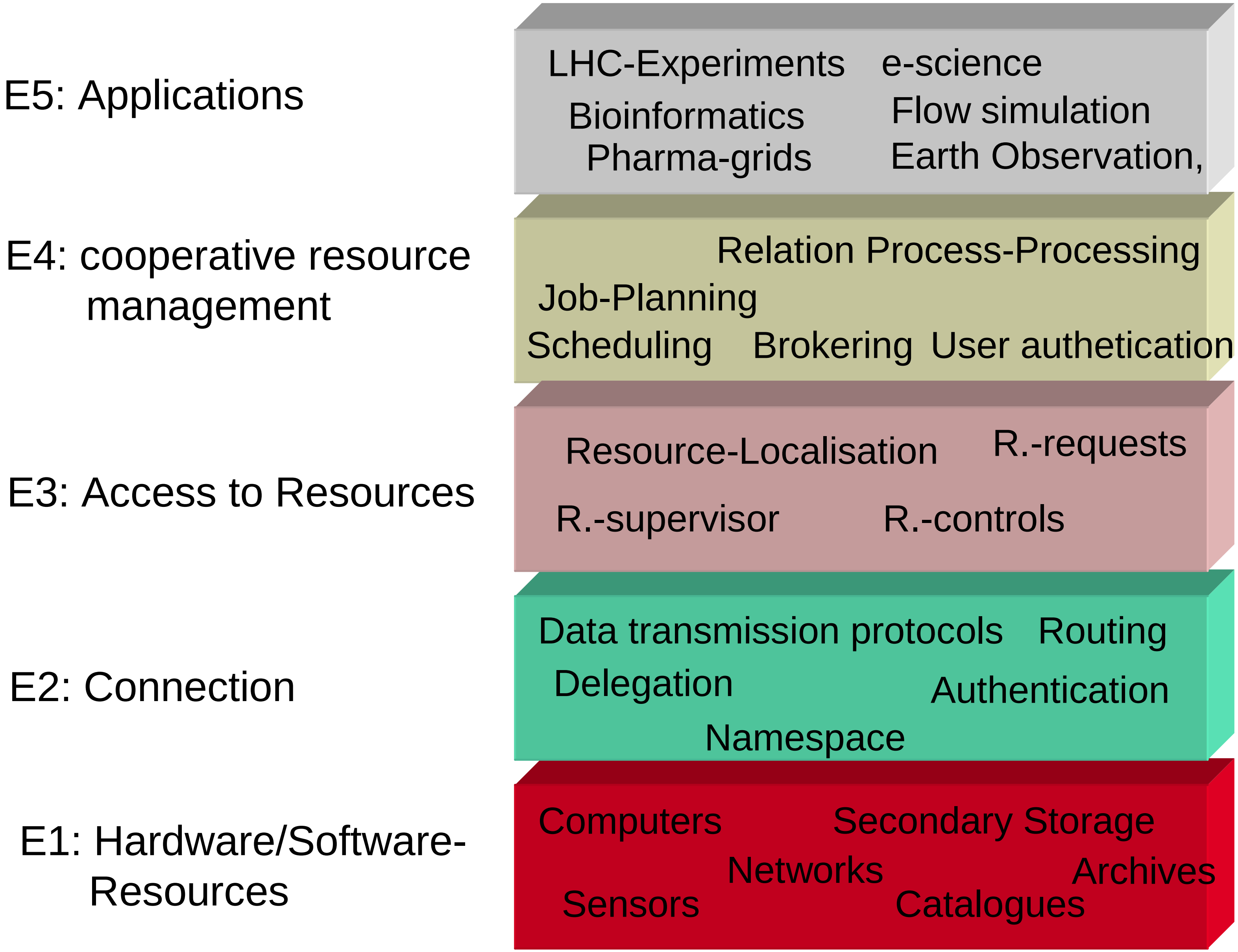


Hoffmann, Reinefeld, Putzer

The “One-Stop Shopping” view of the GRID

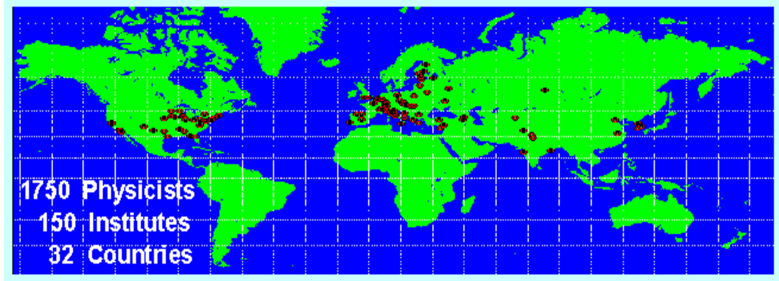


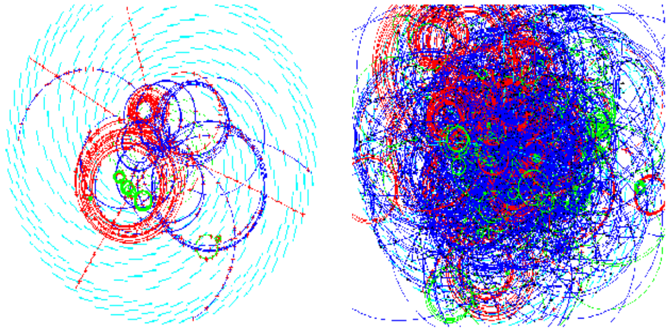
To make European or global Infrastructure: **need use cases, adapt middleware**


E5: Applications
LHC-Experiments
e-science
Bioinformatics
Flow simulation
Pharma-grids
Earth Observation,
E4: cooperative resource management
Relation Process-Processing
Job-Planning
Scheduling
Brokering
User authetication
E3: Access to Resources
Resource-Localisation
R.-requests
R.-supervisor
R.-controls
E2: Connection
Data transmission protocols
Routing
Delegation
Authentication
Namespace
E1: Hardware/Software-Resources
Computers
Secondary Storage
Networks
Archives
Sensors
Catalogues

The LHC Computing Challenge

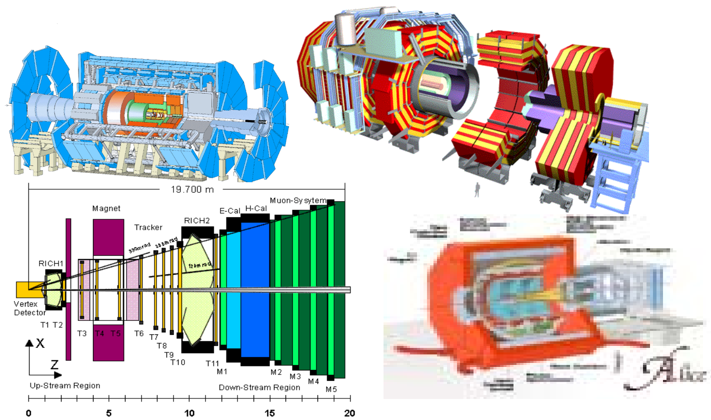
- **Geographical Dispersion** of people and resources
- Communication and collaboration at a distance
- Distributed computing resources
- Remote software development and physics analysis





- **Complexity** of the detectors and LHC environment
- 1,000,000,000 highly selected events/year

- **Scale** of the data to be treated
- Approx. 8 PetaBytes of data/year (10 million CD-ROMS)
- Need 200,000 of today's PCs to process the data



Meeting the challenge: **the LCG project**

LHC Computing project creating a

Global Virtual Computing Centre for Particle Physics

Goal - prepare and deploy the LHC Computing environment

applications - tools, frameworks, environment, persistency

computing system → global grid service

- cluster → automated fabric
- collaborating computer centres → grid
- CERN-centric analysis → global analysis environment

The Grid Vision

Researchers perform their activities around the world, interact with colleagues, share, access data and applications

The GRID:
networked processing centres; ”middleware” (software): the “glue” of resources; Provides Compute power

Scientific instruments, industrial applications provide huge amount of data

Virtual distributed Computing Centre

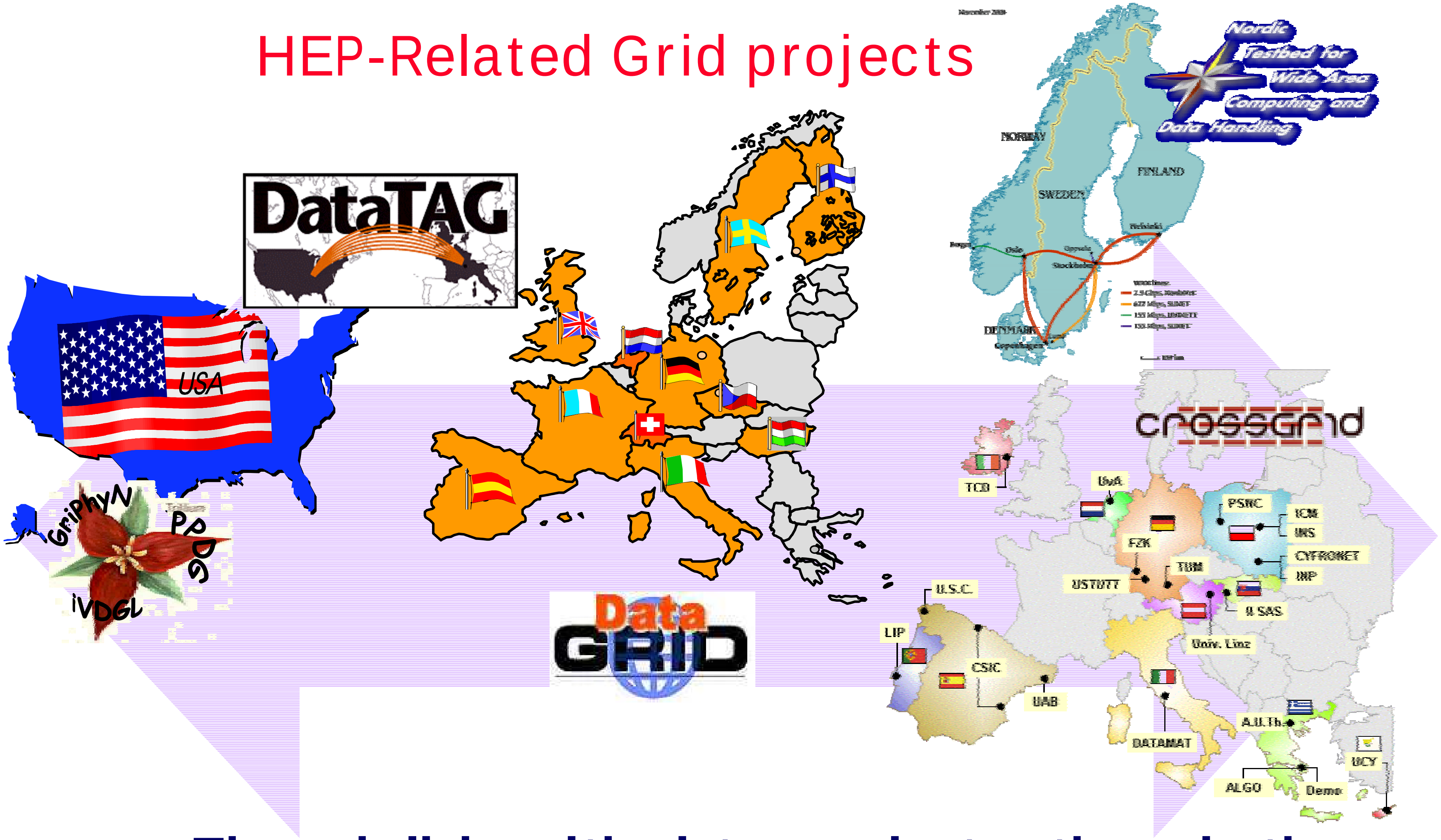
The user ---

sees the a "utility" on his/her laptop

does not need to know - where the data is
- where the processing capacity is
- how things are interconnected
- the details of the different hardware

and is not concerned by the conflicting policies of the individual centres

HEP-Related Grid projects



Through links with sister projects, there is the potential for a truely global scientific applications grid

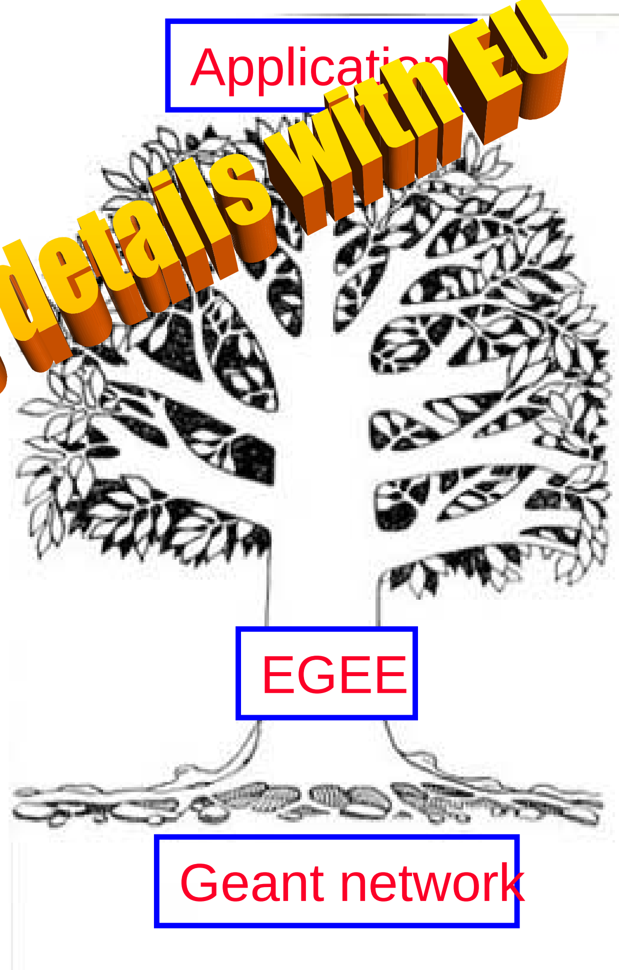
EGEE vision

Enabling Grids for E-science in Europe

(proposal submitted to EU, FP6 I3, in May 2003)

- 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
  - International connections (US and AP)
  - Several pioneering prototype results
  - Large Grid re-engineering/deve[illegible]n
  - Goal can be achieved f[illegible] 4 years on top of the natio[illegible]al initiatives
- Approach
  - L[illegible]d planned national and re[illegible]programmes (e.g. LCG)
  - W[illegible]sely with relevant industrial Grid developers, NRNs and US-AP projects

EGEE is invited to negotiate details with EU



Grid Deployment - LCG1

LCG-1 Service

- Certification and distribution process established and tested at ten sites
- Middleware package under test – components from European DataGrid (EDG) and the US grid projects toolkit (VDT)
- Agreement reached on initial principles for registration and security
- RAL to provide the initial grid operations centre
- FZK to provide the call centre
- Target date for opening the service – 1 July

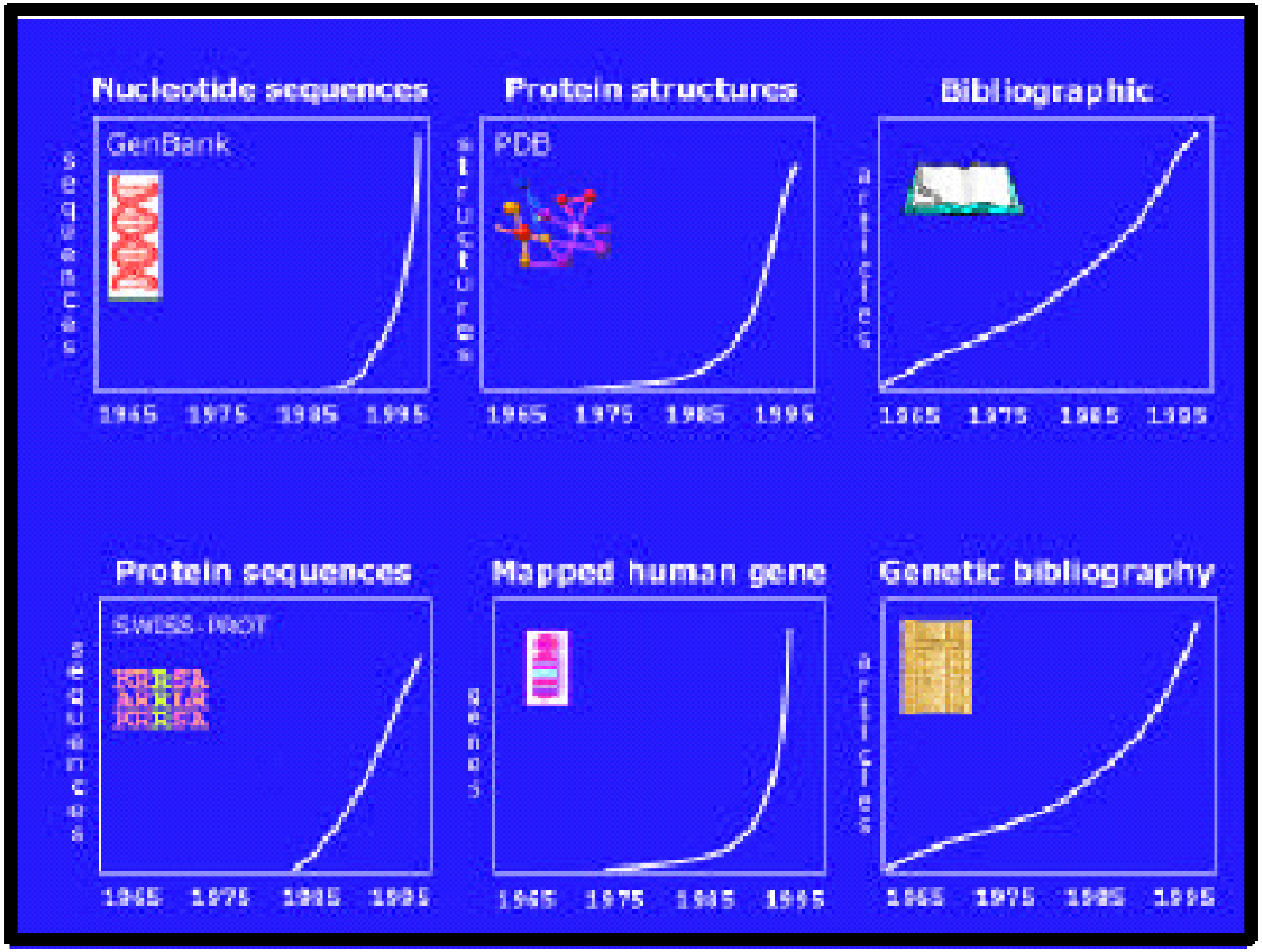
Centres taking part in the LCG prototype service (2003-05)

around the world → around the clock

Conclusions

Challenge and opportunity: Data Explosion

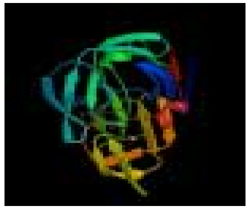
discovery relevant data double every 6-9 months

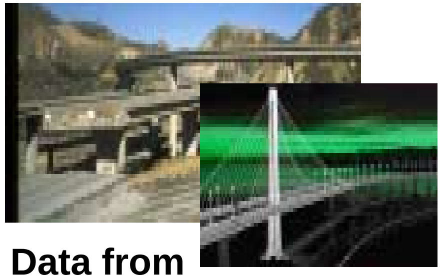


Source: NCBI

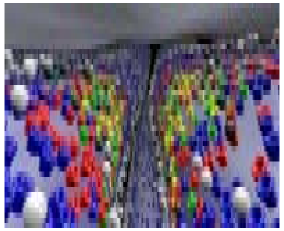
Data: A "Killer App" for the Grid

- Over the next decade, data will come from everywhere
  - **Scientific instruments**
  - **Experiments**
  - **Sensors and sensor nets**
  - **New devices (personal digital devices, computer-enabled clothing, cars, …)**
- And be used by everyone
  - **Scientists**
  - **Consumers**
  - **Educators**
  - **General public**
- SW environment will need to support unprecedented diversity, globalization, integration, scale, and use

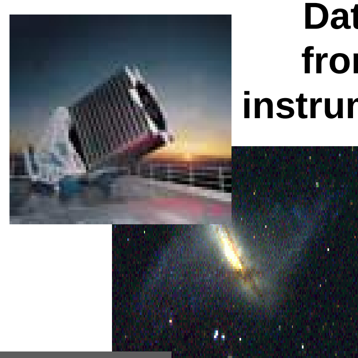




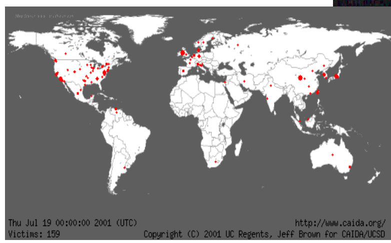
Data from sensors



Data from simulations



Data from instruments



Data from analysis

from Rich Hirch, NSF

The Information Tsunami

Terabyte [10^{12} bytes]

1 Terabyte: An automated tape robot OR all the X-ray films in a large technological hospital OR 50000 trees made into paper and printed OR daily rate of EOS data (1998);

2 Terabytes: An academic research library OR a cabinet full of Exabyte tapes;

10 Terabytes: The printed collection of the US Library of Congress;

50 Terabytes: The contents of a large Mass Storage System;

400 Terabytes: National Climactic Data Center (NOAA) database;

Petabyte [10^{15} bytes]

1 Petabyte: 3 years of EOS data, OR 1 sec of CMS data collection

2 Petabytes: All US academic research libraries

8 Petabytes: All information available on the Web;

20 Petabytes: Production of hard-disk drives in 1995;

200 Petabytes: All printed material OR production of digital magnetic tape in 1995;

Exabyte [10^{18} bytes]

2 Exabytes: Total volume of information generated worldwide annually.

5 Exabytes: All words ever spoken by human beings.

Zettabyte [10^{21} bytes]

Yottabyte [10^{24} bytes]

from Rich Hirch, NSF

Five Principles for Cyber Infrastructure

adapted from a NSF, DOE, CERN discussion

[0] The cost and complexity of 21st Century Science requires the creation of advanced and coherent global Information Infrastructure (Infostructure).	**(global cost sharing)**
[1] The construction of a coherent Global Infostructure for Science require definition and drivers from Global Applications (that will also communicate with each other)	**(applications give benefits that justify expense)**
[2] Further, forefront Information Technology must be incorporated into this Global Infostructure for the Applications to reach their full potential for changing the way science is done..	**(Frontiers & Global for real changes)**
[3] LHC is a near term Global Application requiring advanced and un-invented Infostructure and is ahead in planning compared to many others.	**(LHC is a frontier)**
[4] U.S. agencies must work together for effective U.S. participation on Global scale infostructure and the successful execution of the LHC program in a 4 way agency partnership, with international cooperation in view.	**(national then global)**

From Scientists to virtual Organisations

For computing, common perception is often one PC and all the services and software on this single PC.

Going from a software running locally with plenty of "do-it yourself" fixes to a "full-proof, ready to install everywhere" product is not an easy task, but no magic either

nor to use growing experience of Grid computing

However:

Several cultural and technical problems need to be addressed :

private - public

proprietary - open

text mining - information extraction

data challenge – knowledge creation

Particle Physics solves this by

Open Virtual Global Collaborations

What will you do??