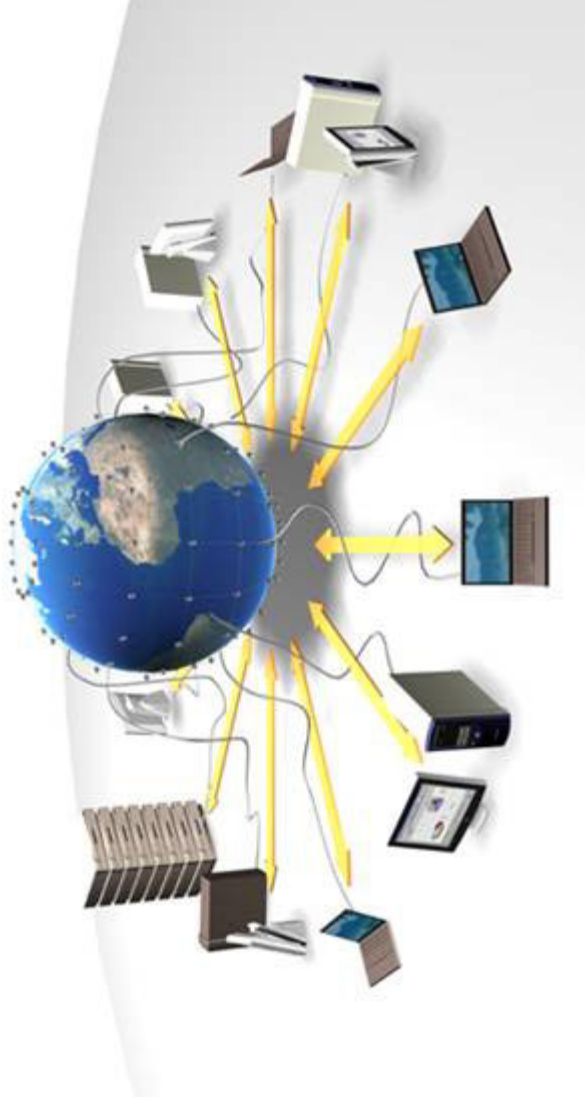


Why CERN provides a unique IT challenge for industry: from the Web to the Grid

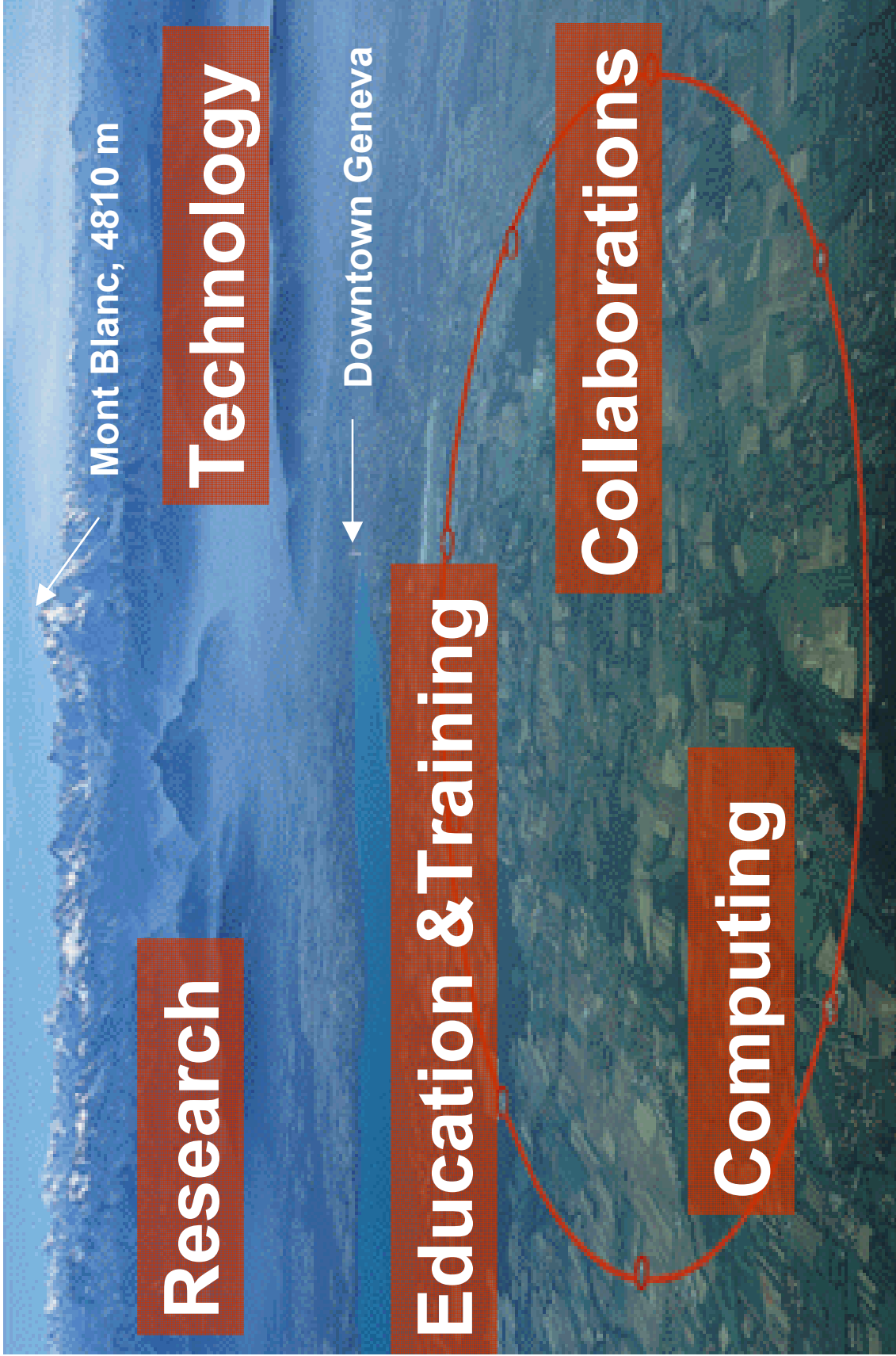


*Hans Hoffmann, Director Technology Transfer
& Scientific Computing, (1999-2003)*





CERN



Mont Blanc, 4810 m

Research

Technology

Downtown Geneva

Education & Training

Collaborations

Computing



Overview

- **What is CERN, What is the LHC? Why particle physics creates a major challenge for IT.**
- **What is the Grid at CERN? An overview of major Grid projects at CERN, and their industrial perspectives.**
- **The IT Department at CERN. The background for the rest of today's programme.**





What is CERN?

CERN is:

- Some **2500 staff scientists** (physicists, engineers, ...)
- Some **6500 visiting scientists** (half of the world's particle physicists)

They come from 500 universities representing 80 nationalities.

- CERN is the world's largest particle physics centre
- Particle physics is about:
 - elementary particles which all matter in the Universe is made of
 - fundamental forces which hold matter together
- Particle physics requires:
 - special tools to create and study new particles



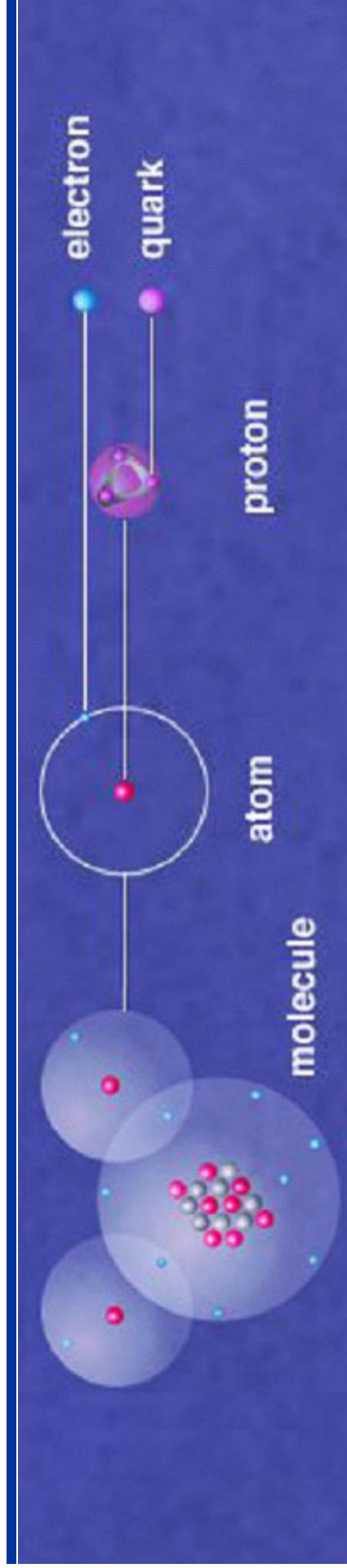


What is CERN?

- Physicists smash particles into each other to:
 - identify their **components**
 - **create** new particles
 - reveal the nature of the **interactions** between them
 - create an environment similar to the one present at the origin of our Universe

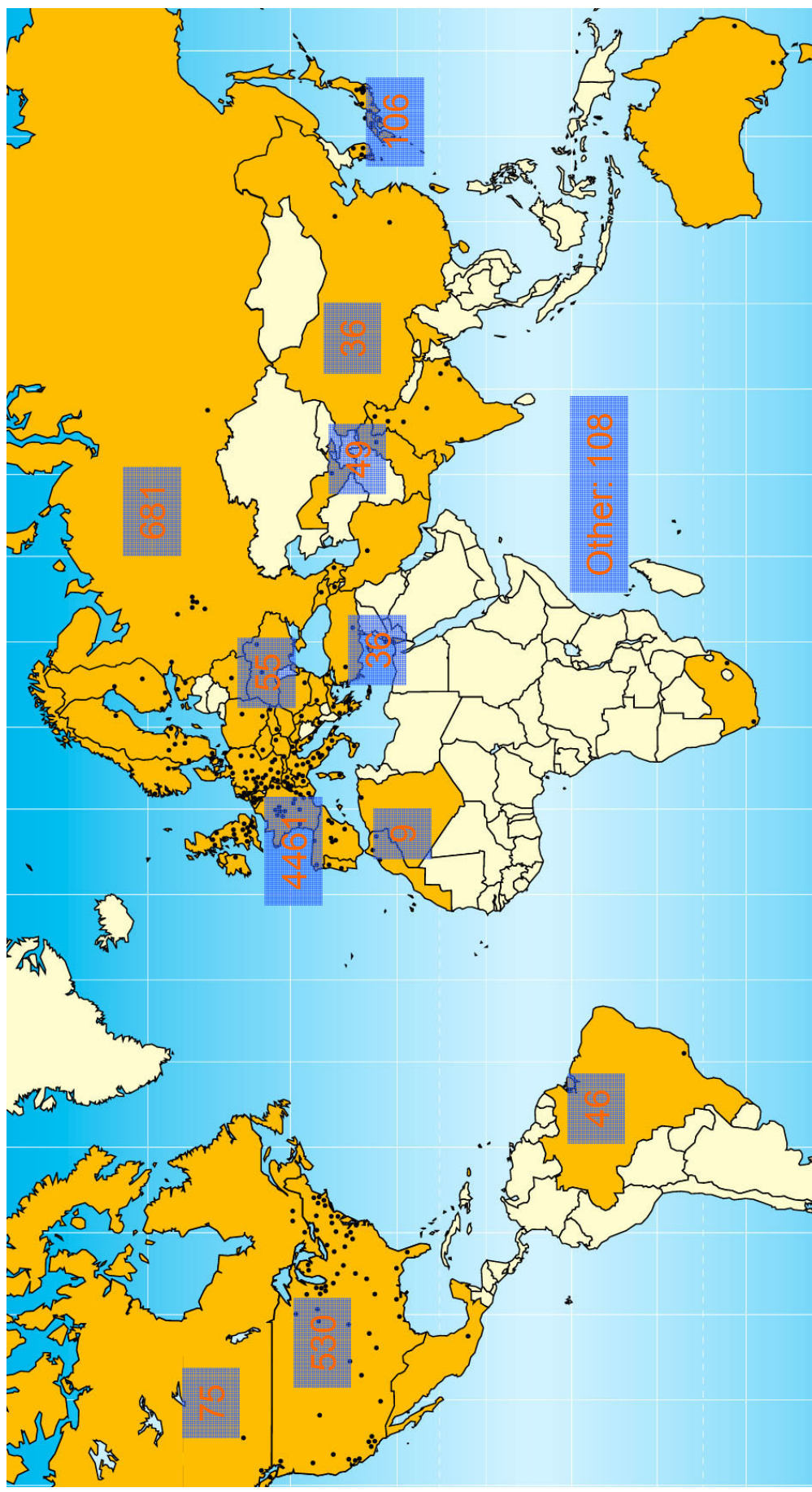


- **What for?** To answer fundamental questions like:
how did the Universe begin? What is the origin of mass?
What is the nature of antimatter?





CERN & Collaborating Institutes





What is the LHC?

- LHC will collide beams of protons at an energy of 14 TeV
- Using the latest super-conducting technologies, it will operate at about -270°C , just above absolute zero of temperature.
- With its 27 km circumference, the accelerator will be the largest superconducting installation in the world.

LHC is due to switch on in 2007

Four experiments, with detectors as 'big as cathedrals':

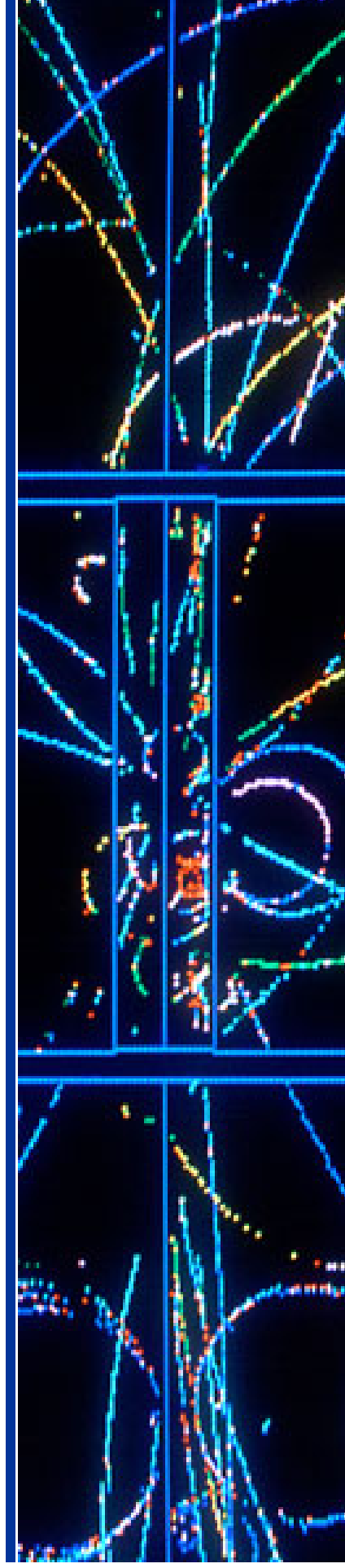
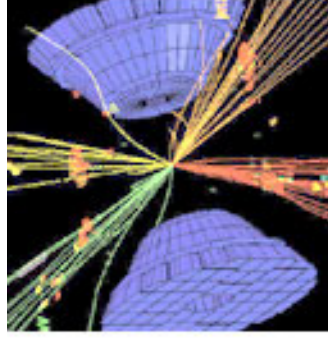
ALICE
ATLAS
CMS
LHCb





The LHC Data Challenge

- A particle collision = an event
- Physicist's goal is to count, trace and characterize all the particles produced and fully reconstruct the process.
- Among all tracks, the presence of “special shapes” is the sign for the occurrence of interesting interactions.





LHC data

- 40 * 25 million collisions per second
- After filtering, 100 collisions of interest per second
- > 1 Megabyte of data digitised per collision = recording rate up to 1 Gigabyte/sec
- 10^{10} collisions recorded each year = 10 Petabytes/year of data

1 Megabyte (1MB)
A digital photo

1 Gigabyte (1GB)
= 1000MB

5GB = A DVD movie

1 Terabyte (1TB)
= 1000GB

World annual book production

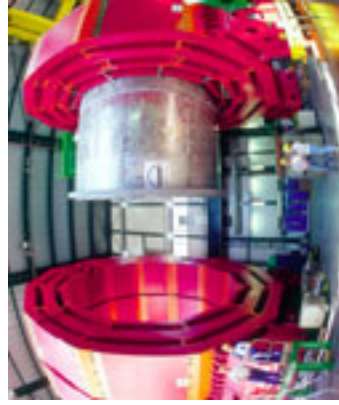
1 Petabyte (1PB)
= 1000TB

Annual production of one LHC experiment

1 Exabyte (1EB)
= 1000 PB

3EB = World annual information production

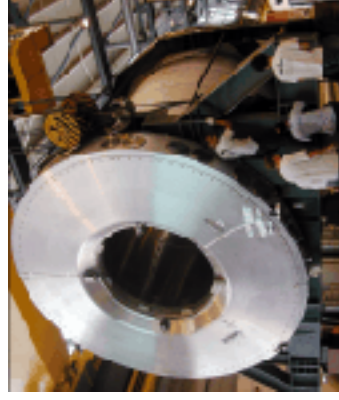
CMS



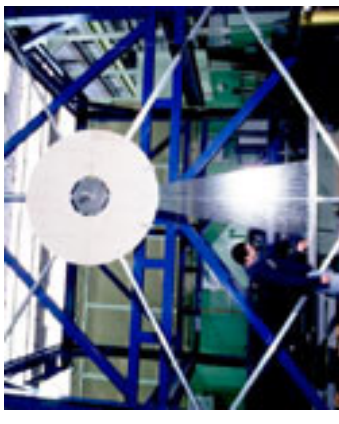
LHCb



ATLAS



ALICE

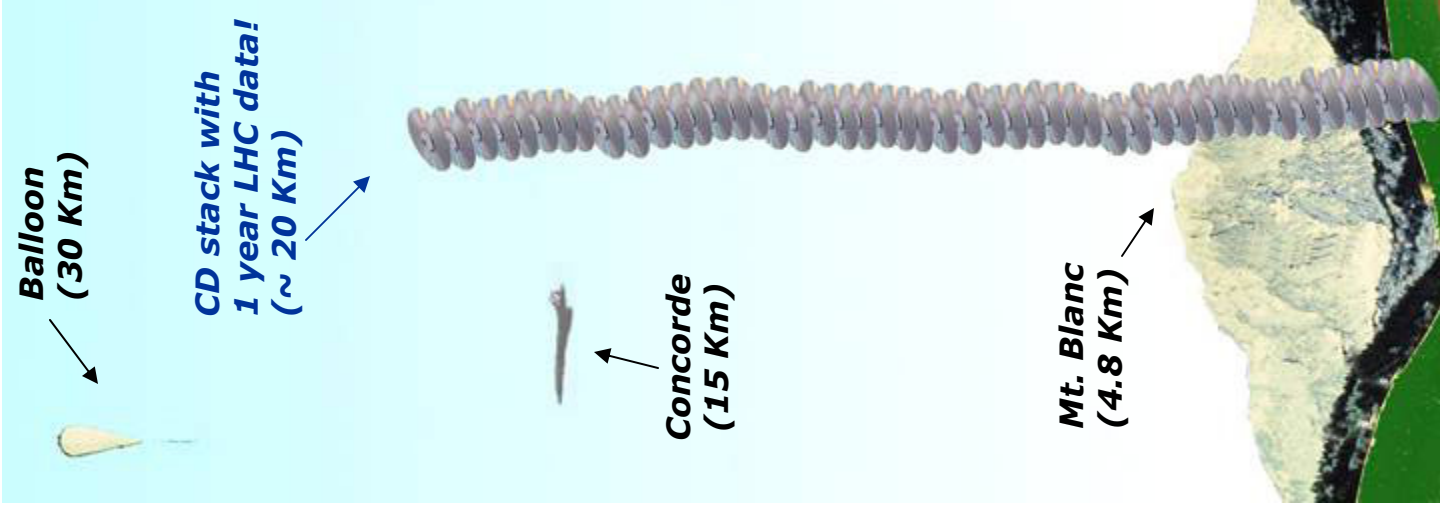




LHC data

**LHC data correspond to about
20 million CDs each year**

**Where will the
experiments store all of
these data?**

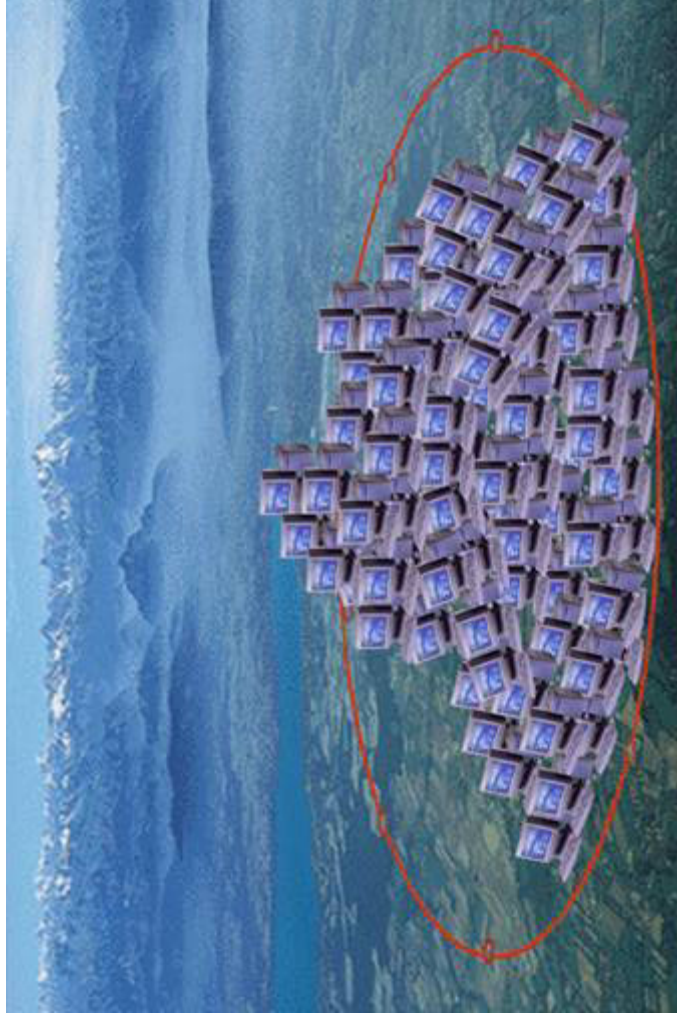




LHC processing

LHC data analysis requires a computing power equivalent to ~ 100,000 of today's fastest PC processors

Where will the experiments find such a computing power?





Computing at CERN

- High-throughput computing based on reliable “commodity” technology
- More than 1000 dual processor PCs
- More than 1 Petabyte of data on disk and tapes
- Largest PC farm in Europe

Nowhere near enough!





Computing at CERN



CPU servers

The new computer room in the vault of the Computer Centre is being populated...



Disk servers



Tape silos and servers

NB: Major tender for offline computing capacity (CPU, storage) in 2005, worth ~60MCHF. Market survey starting 2004. (Tony Cass for details)



Computing at CERN



....the existing computer center is being cleared for renovation....



...and an upgrade of the power supply from 0.5MW to 2.5MW is underway.



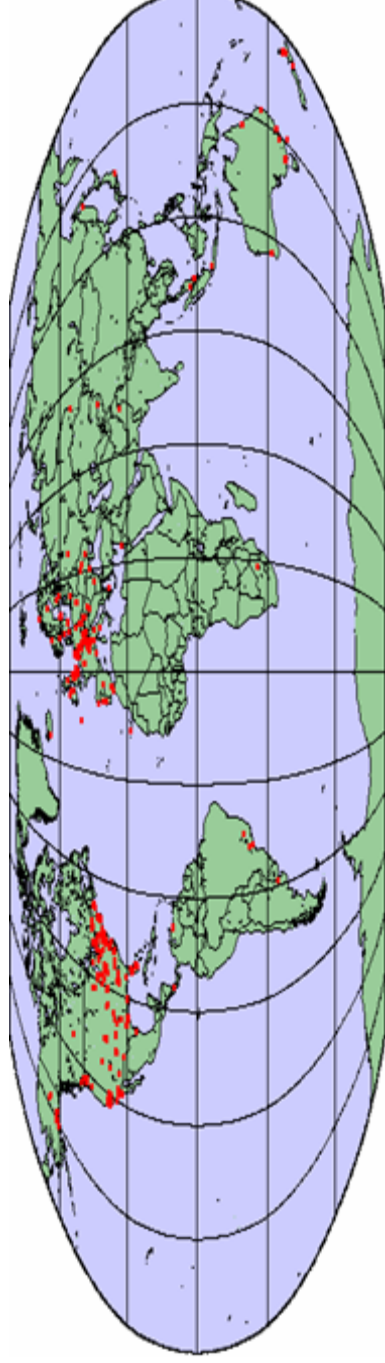
Computing for LHC

Europe:
267 institutes
4603 users

Elsewhere:
208 institutes
1632 users

- **Problem:** even with Computer Centre upgrade, CERN can provide only a fraction of the necessary resources

- **Solution:** Computing centers, which were isolated in the past, will be connected, **uniting the computing resources of particle physicists worldwide**



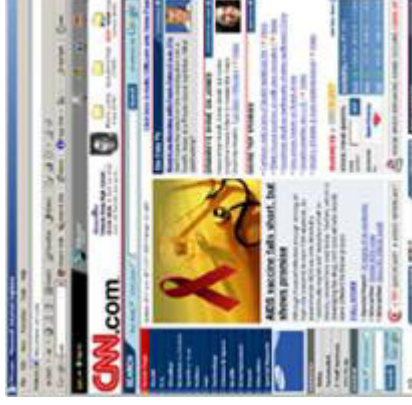


What is the Grid?

- The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations → properly organised, validated: "Next Library of Alexandria"

- In contrast, the **Grid** is an emerging infrastructure that provides seamless access to computing power and data storage capacity distributed at computer centres around the globe.

NB: the Grid is **much more complex** than distributed clusters within one organisation or cycle scavenging from PCs (SETI@home).





What is the Grid?

From an industry point of view:

- The Grid is, for the time being, many Grids
- Grids are a framework for Resource Virtualisation
- Grids are a label for Wide-Area Distributed Computing
- Grids will rely on new models of Inter-Organisational Security
- Grids will be used by Virtual Organisations
- Grids are a platform for Resource Discovery

Grid Services will be based on Web Services

Grid development has been driven by the academic community
Industrial variants are on-demand computing, Utility computing

Developing the Grid may spin off disruptive technologies



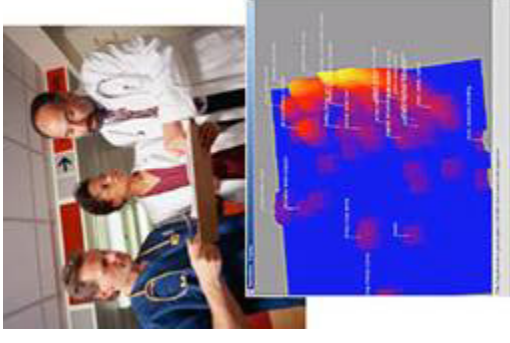
What are the challenges for CERN?

- Must share data between >7000 scientists with multiple interests
 - Must link major computer centres, not just PCs: "one stop"-service
 - Must ensure all data accessible anywhere, anytime
 - Must grow rapidly, yet remain reliable for more than a decade
 - Must cope with different management policies of different centres
 - Must ensure data security: more is at stake than just money!
 - Must be up and running by 2007
-



Grid Applications for Science

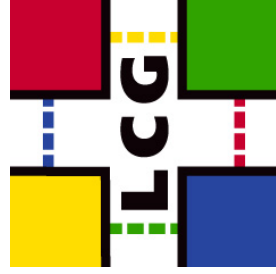
- **Medical/Healthcare** (*imaging, diagnosis and treatment*)
- **Bioinformatics** (*study of the human genome and proteome to understand genetic diseases*)
- **Nanotechnology** (*design of new materials from the molecular scale*)
- **Engineering** (*design optimization, simulation, failure analysis and remote Instrument access and control*)
- **Natural Resources and the Environment** (*weather forecasting, earth observation, modeling and prediction of complex systems*)
- Grid of **Knowledge Metadata** (*combine virtual libraries*)





Grid @ CERN

- CERN projects:
LHC Computing Grid (LCG)
- EC funded projects led by CERN:
Enabling Grids for E-Science in Europe (EGEE)
+others
- Industry funded projects:
CERN openlab for DataGrid applications





LHC Computing Grid (LCG)

Timeline:

- 2002: start project
- 2003: service opened (LCG-1 started in September with 12 sites)
- 2004 LCG-2 released, 40 sites contributing, HP is first industry site.
- 2002 - 2005: deploy the environment for LHC computing
- 2006 – 2008: build and operate the LHC computing service.



Academia Sinica Taiwan, BNL, CERN, CNAF, FNAL, FZK, IN2P3 Lyon, FKI Budapest, Moscow State Uni., Prague, PIC Barcelona, RAL, Uni. Tokyo



See *Tony Cass* for more information



EGEE: Enabling Grids for e-Science in Europe

Mission:

- Deliver 24/7 Grid service to European science; re-engineer and “harden” Grid middleware for production; “market” Grid solutions to different scientific communities
- Be the first international multiscience production Grid facility

Key features:

- 100 million euros/4years, 70 European partner organisations
- first 2-year phase started April 1st 2004
- transfer services to industry in 2nd 2-year phase (Geant model)
- >400 software engineers + service support
- EGEE Industry Consortium established (see www.eu-egee.org)



CERN

openlab for DataGrid applications



Mission:

- Industry-sponsored testing and validation of cutting-edge hardware/software solutions in CERN's Grid environment

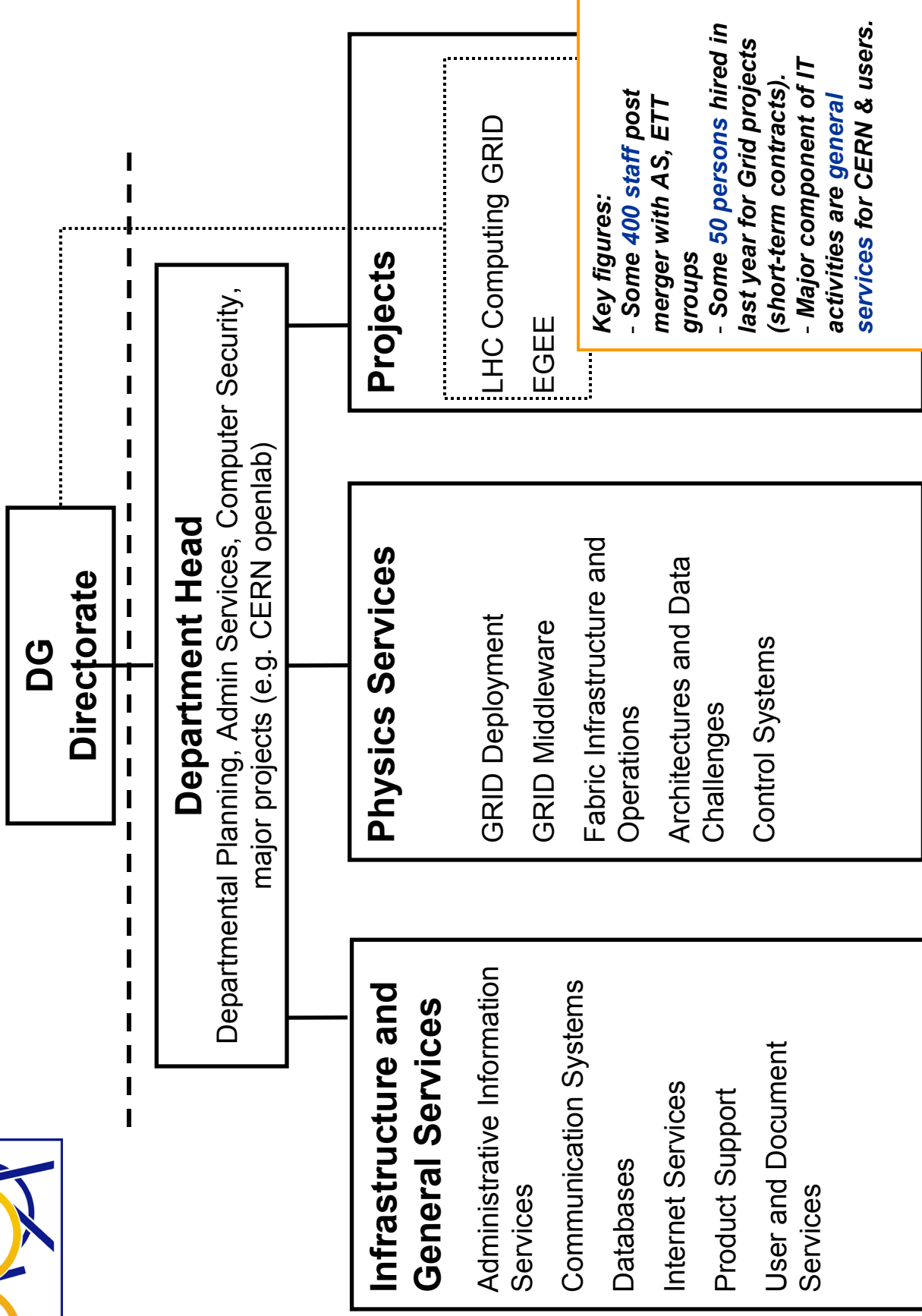
Key features:

- Five partners sponsor 2.5MCHF each over three years (in-kind, sponsorship of postdocs, training, etc.)
- New “contributor” status for smaller firms with promising technologies for LCG (250kCHF sponsorship). Voltaire is first example (joined this month).





New IT Department as of 2004





Conclusions and Perspectives

- LHC data challenge puts CERN at vanguard of Grid technology
- Major opportunities for industry suppliers and technology partners
- Source of computer literate, internationally experienced young people
- IPR for IT at CERN: different models exist (F. Fluckiger talk)
- IT at CERN is much more than just Grid (J. Purvis talk)
- CERN hosting CHEP '04 in Interlaken in September
(<http://chep2004.web.cern.ch/chep2004/>)
- CERN establishing Globe of Innovation with European Industry

