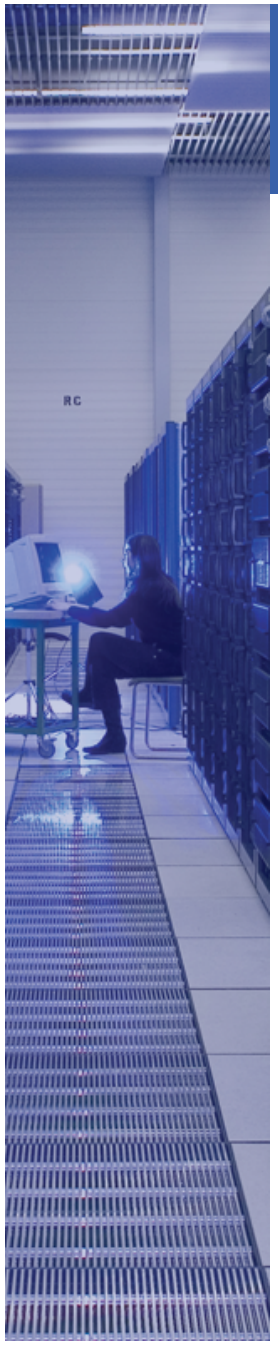
The CERN Computer Centre and the LHC Computing Grid

Università degli Studi di Trieste
Corso di Laurea in Fisica (III anno)

Saturday 9th April 2011

Massimo Lamanna
IT Data Storage Services

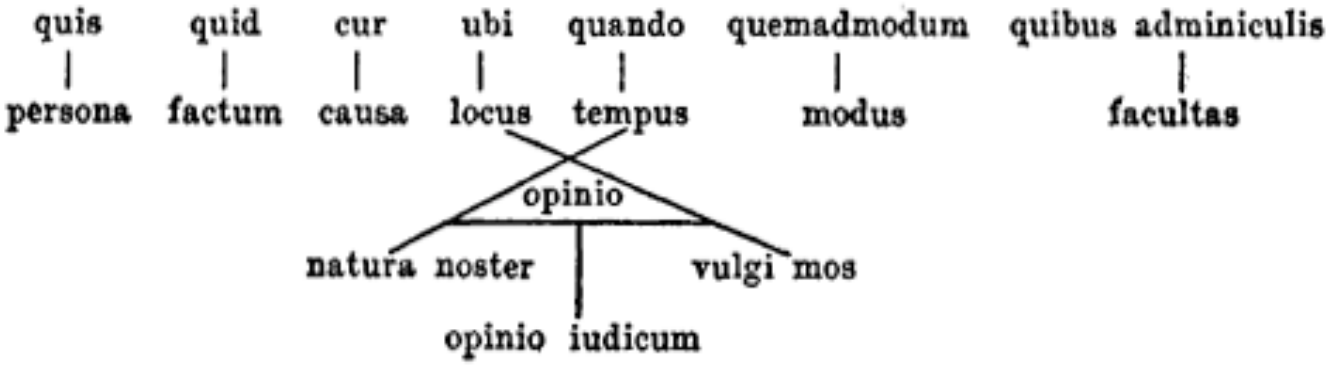


High-Energy Physics Computing

CERN IT Department

- *I keep six honest serving-men*
  (They taught me all I knew);
  Their names are What and Why and When
  And How and Where and Who

» *R. Kipling*



» *Cicerone et al.*

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

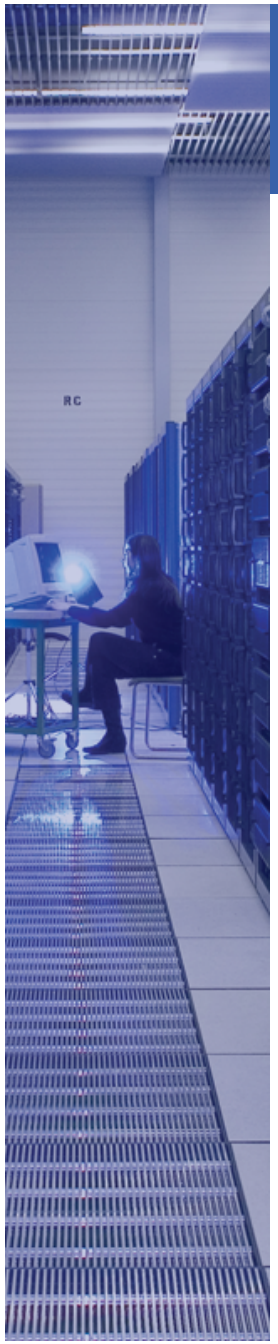
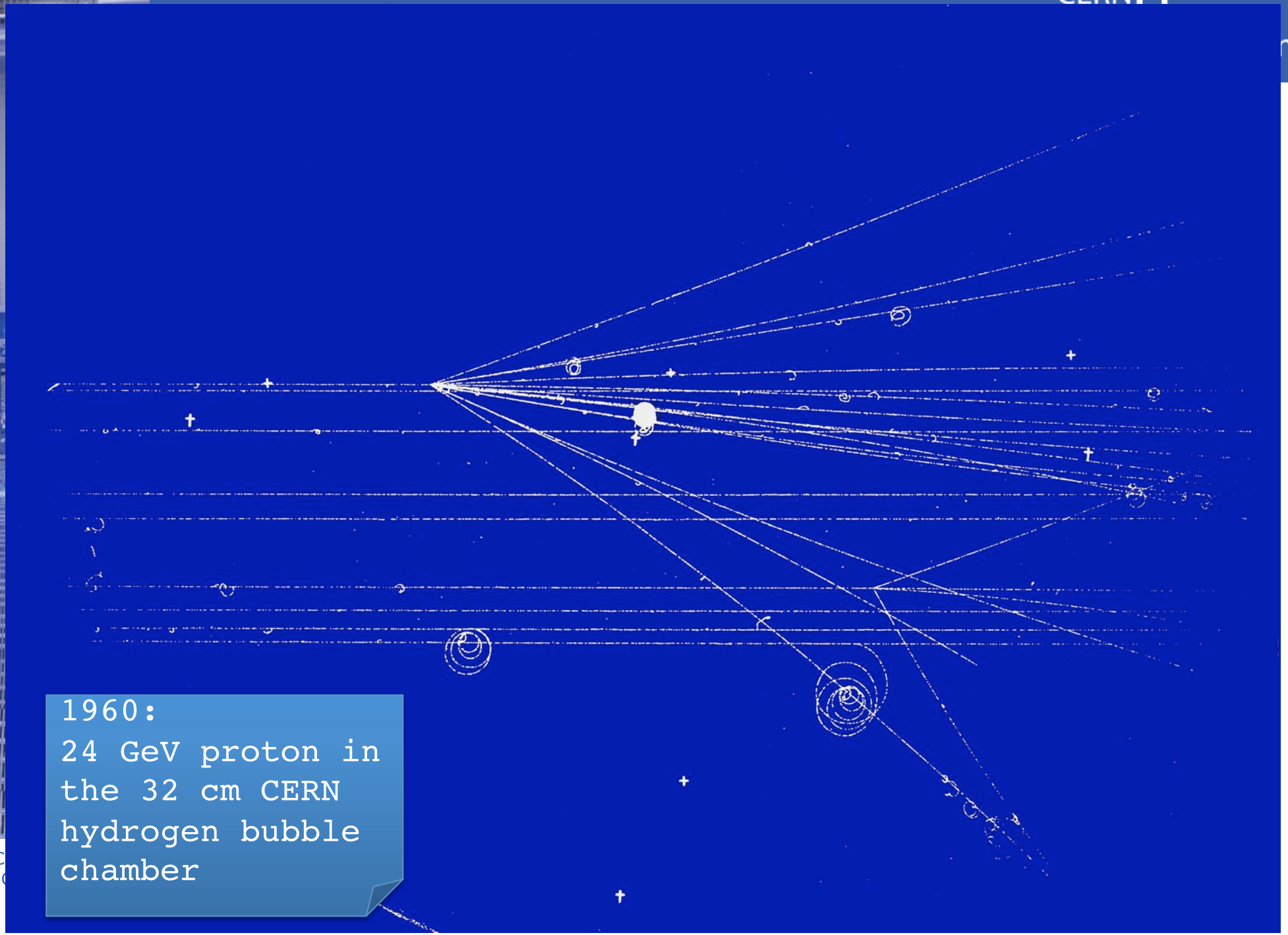


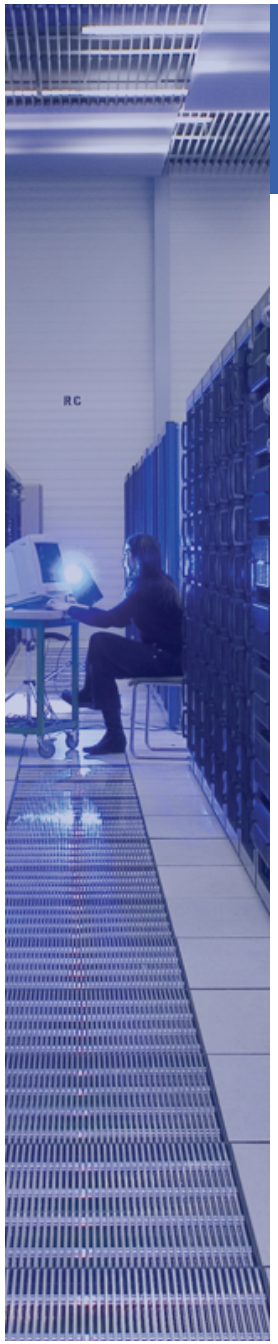
Table of content

- What
- Why
- When
- How
- Where
- Who

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

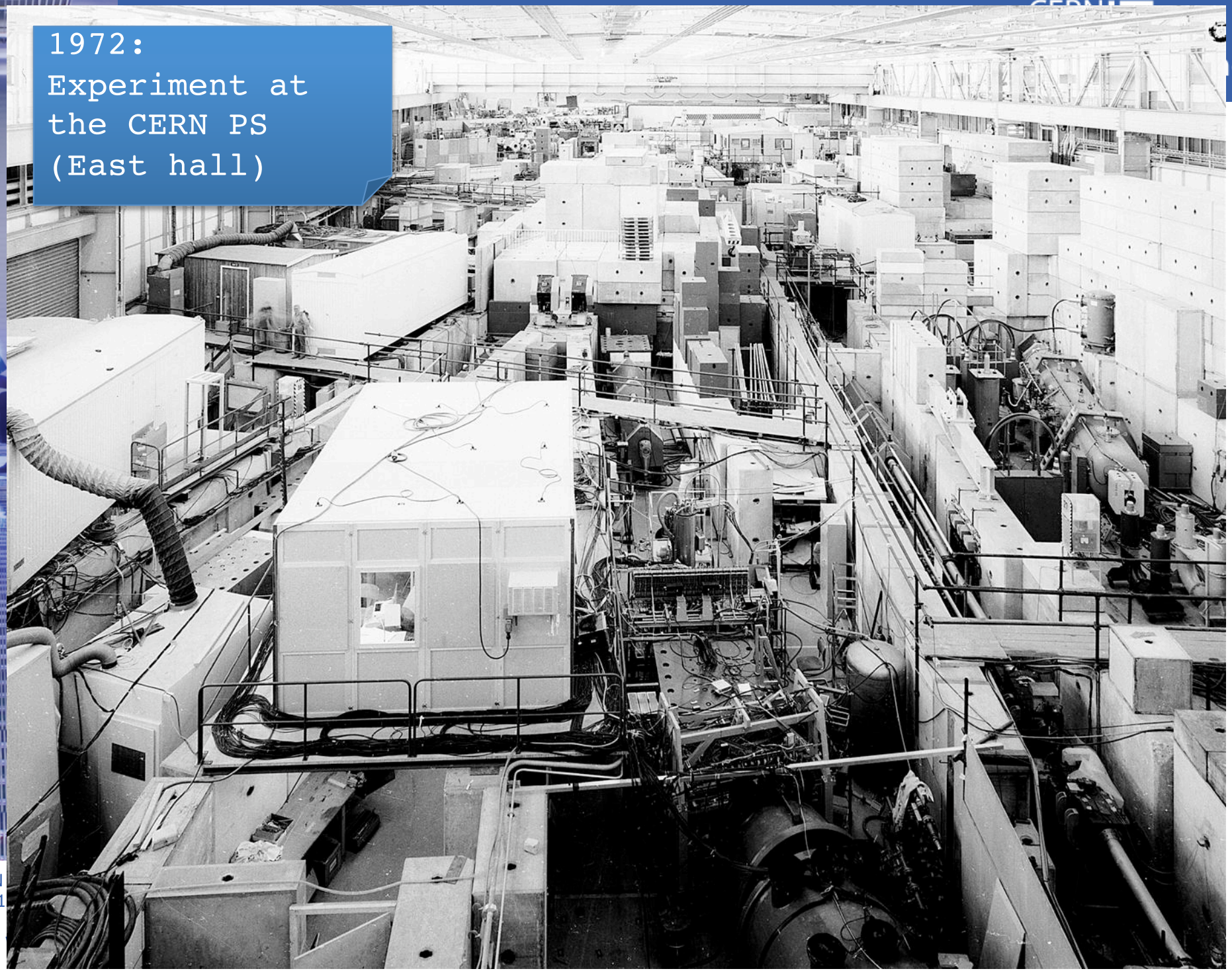


1960:
24 GeV proton in the 32 cm CERN hydrogen bubble chamber



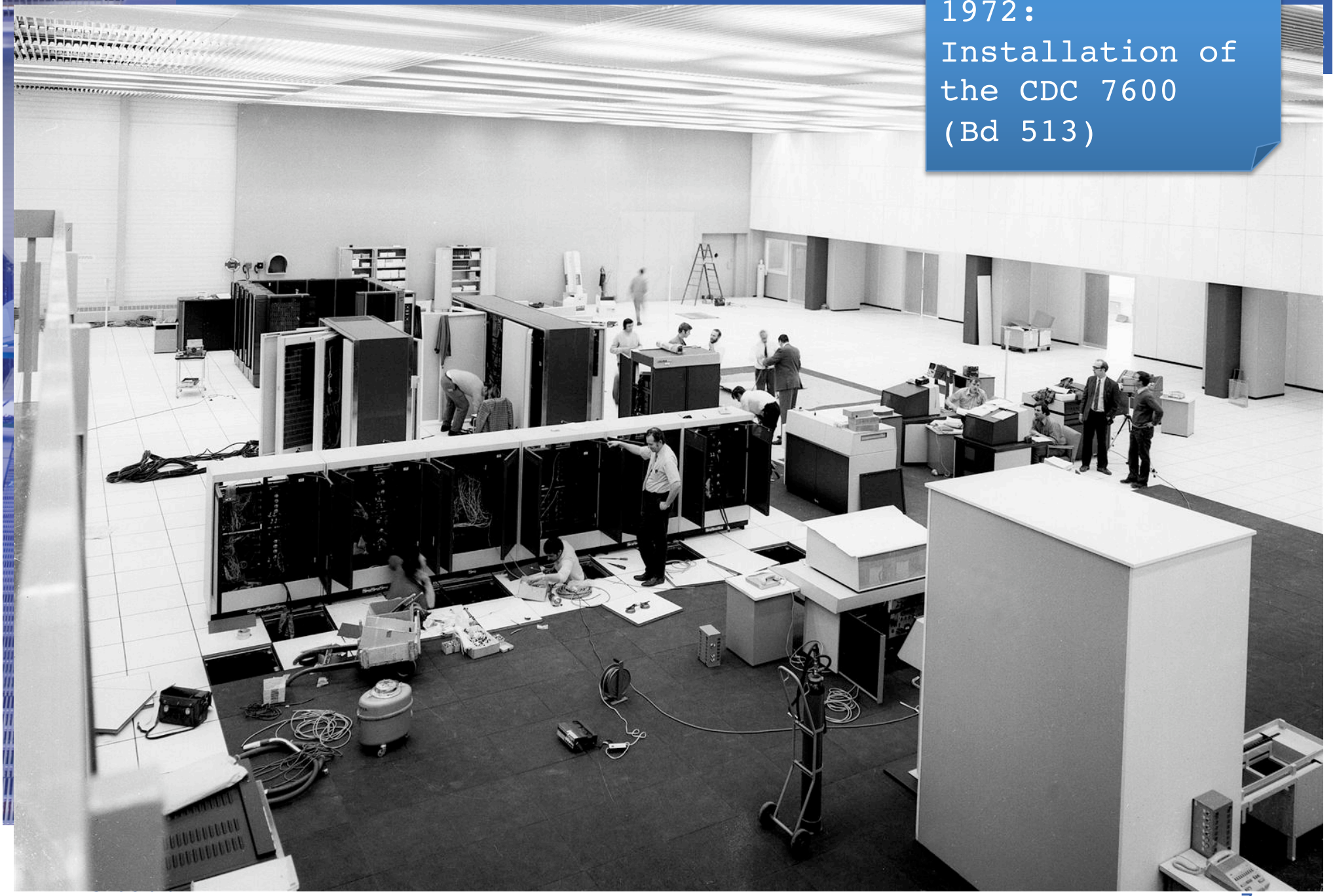
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

1960:
IBM 709 at
Geneva airport

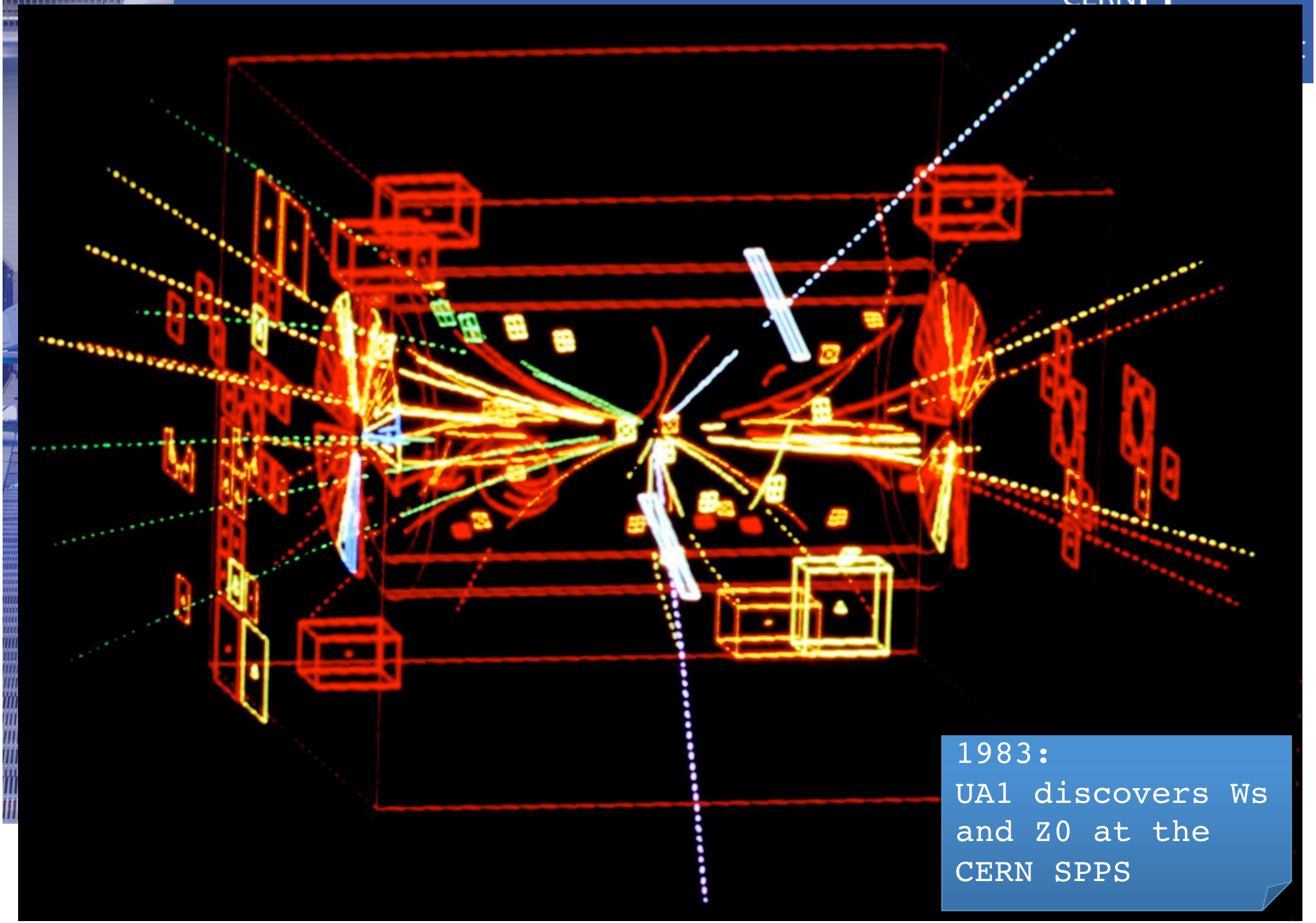


1972:
Experiment at
the CERN PS
(East hall)

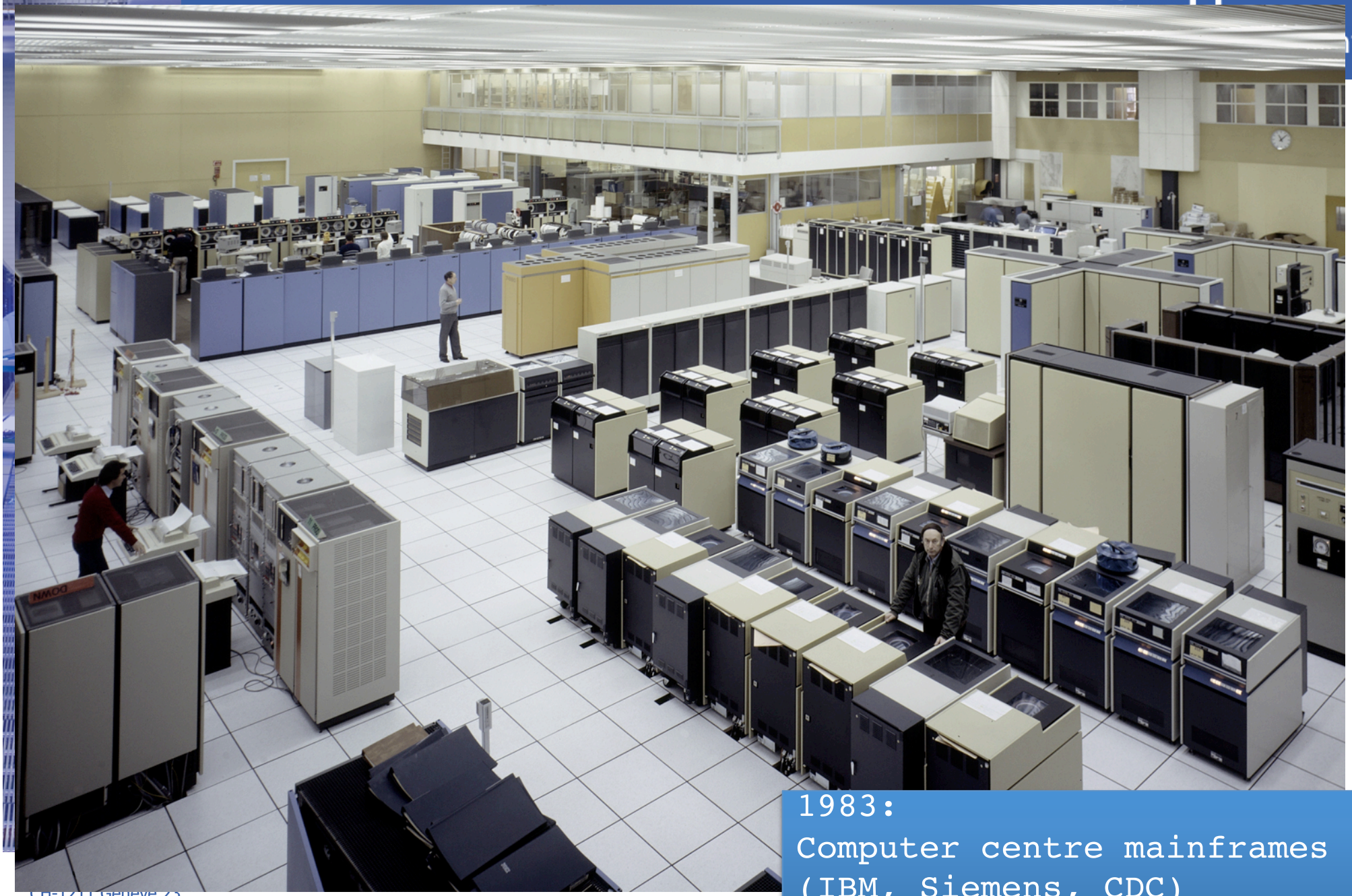
CERN
CH-1



1972:
Installation of the CDC 7600 (Bd 513)

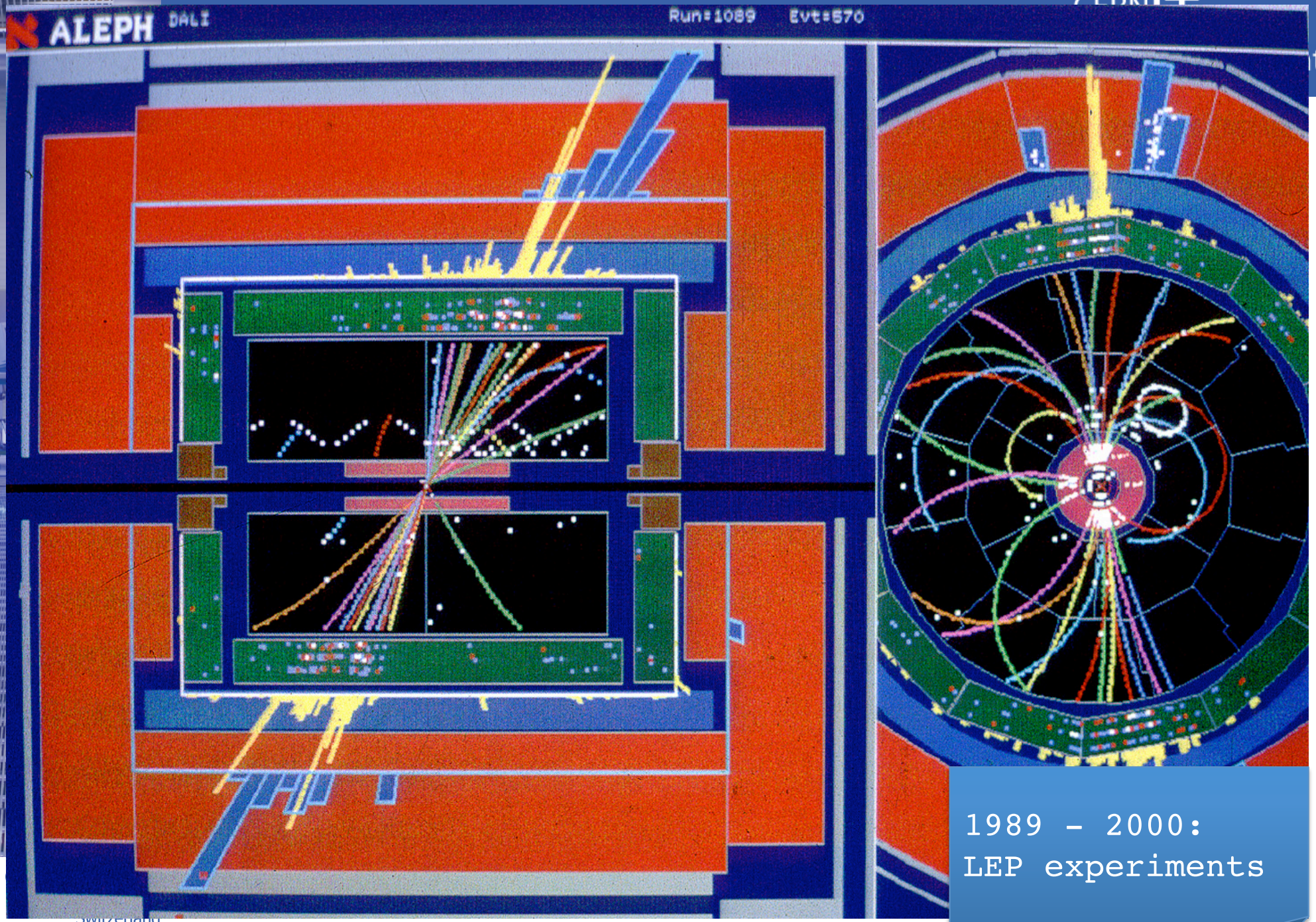


1983:
UA1 discovers Ws and Z0 at the CERN SPPS



1983:
Computer centre mainframes
(IBM, Siemens, CDC)
(Bd 513)

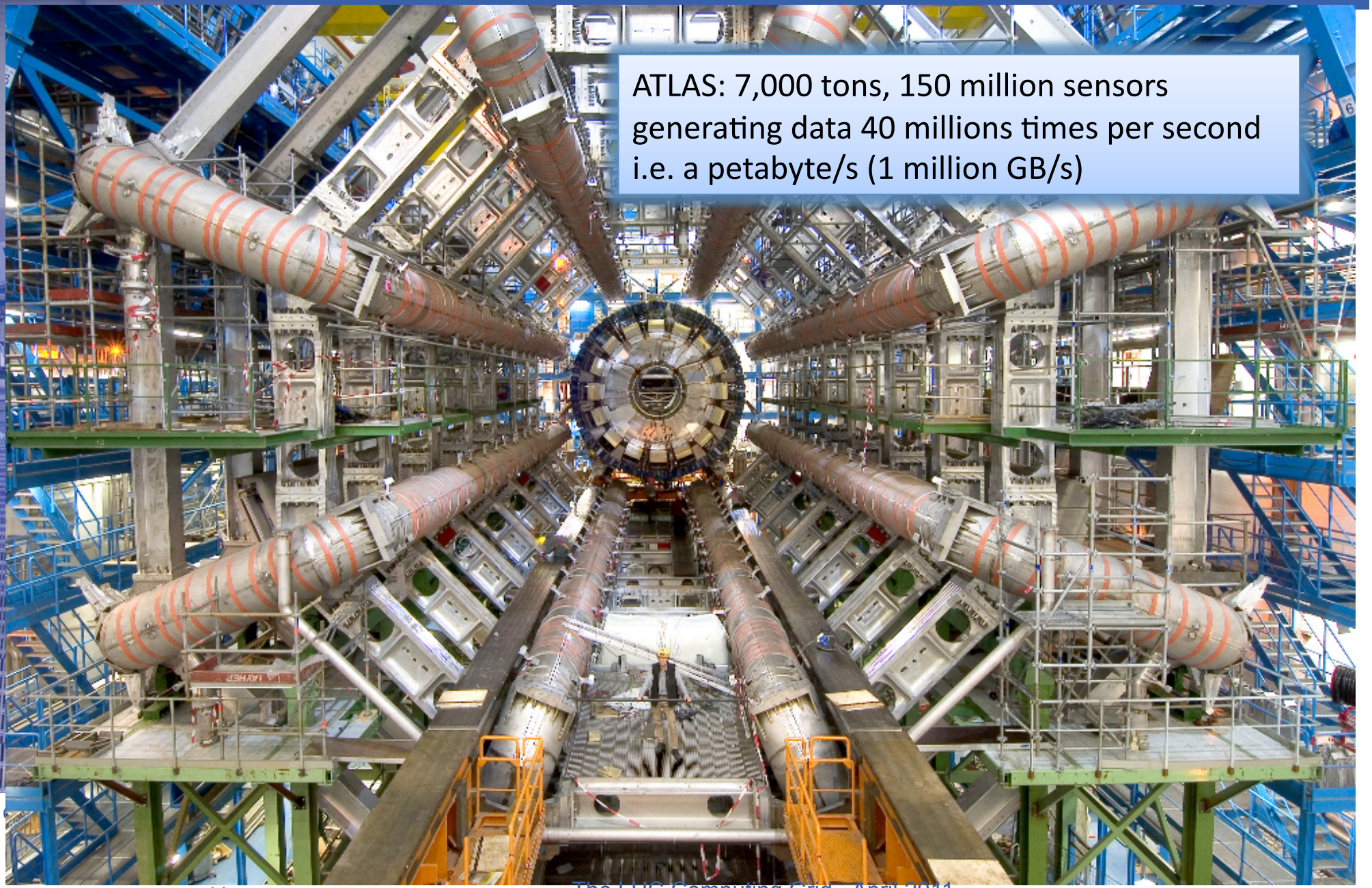
CH-1211 Geneve 23
Switzerland
www.cern.ch/it



1989 - 2000:
LEP experiments

LHC experiments (ATLAS)

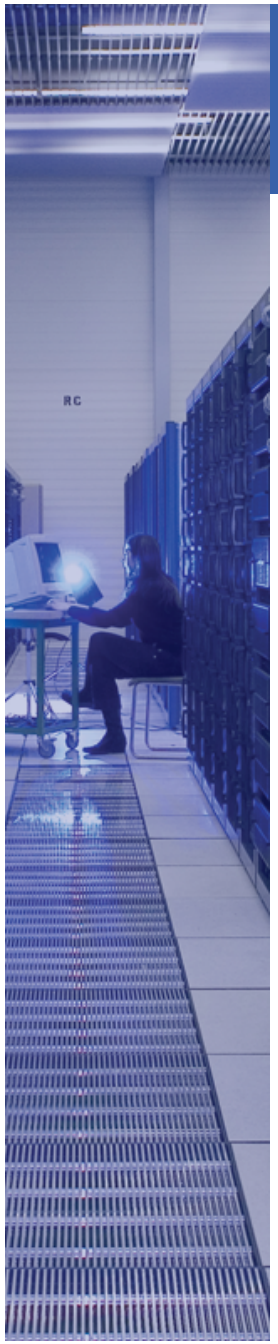
CERN IT Department



ATLAS: 7,000 tons, 150 million sensors generating data 40 millions times per second i.e. a petabyte/s (1 million GB/s)

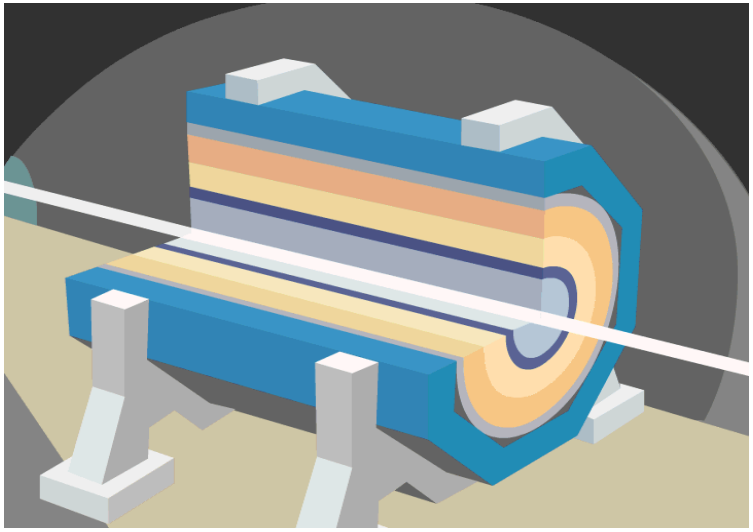
Another view of ATLAS

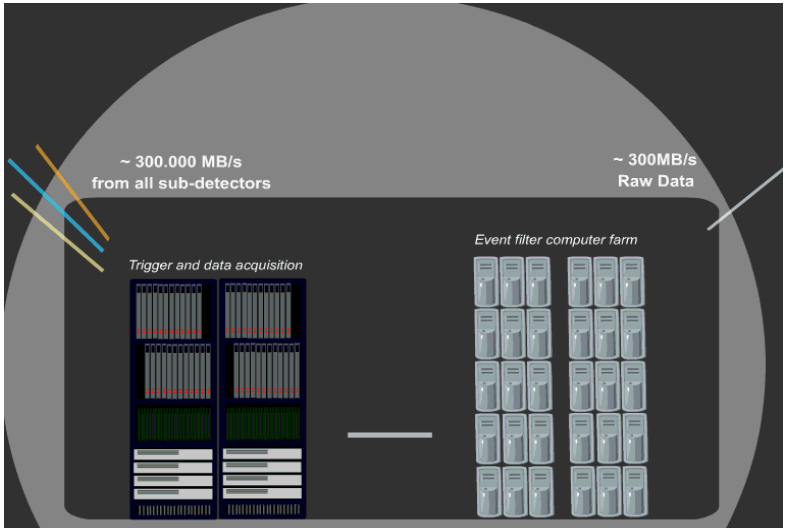
ATLAS is around than 3,000 collaborators
From 169 universities from 37 countries
~1000 students!!!

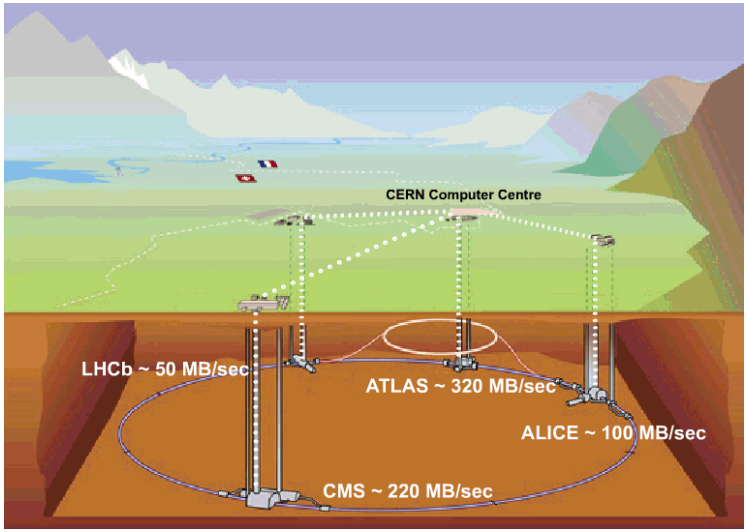


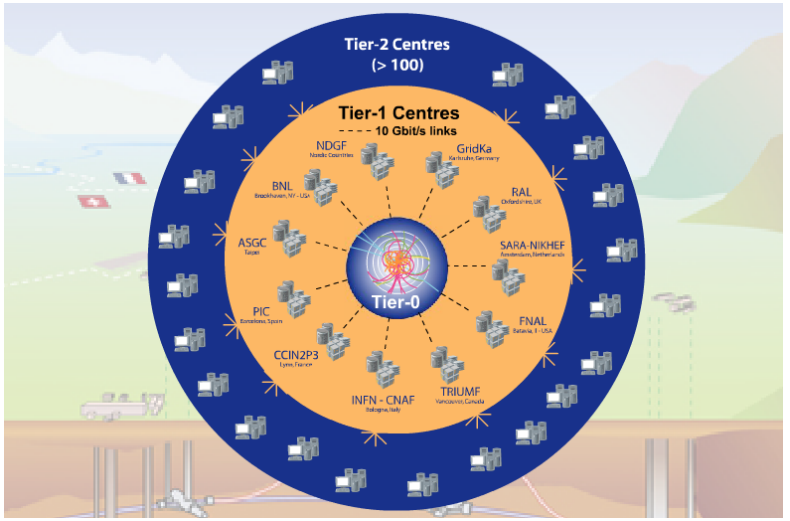
The role of the CERN Computer Centre

CERN IT Department

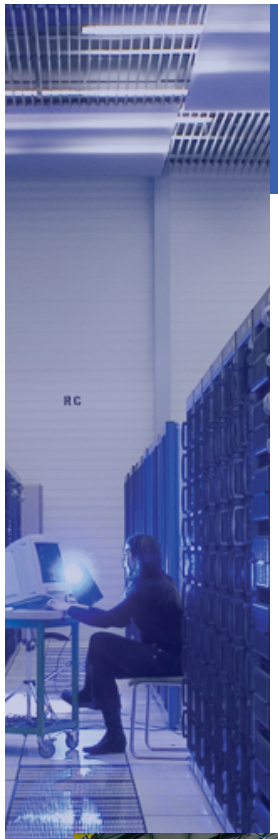






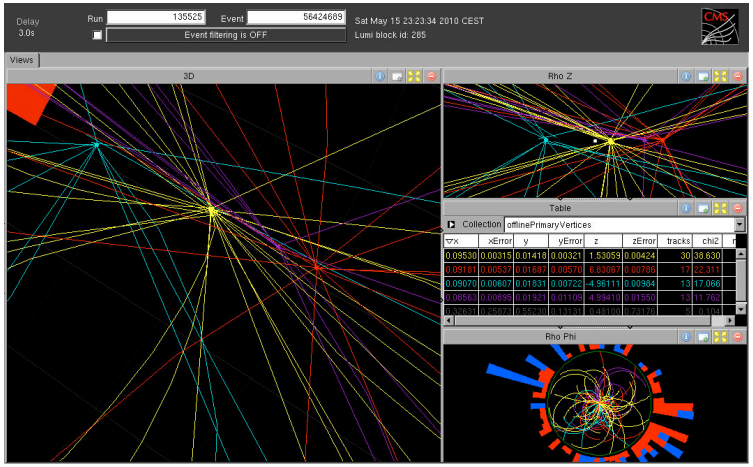


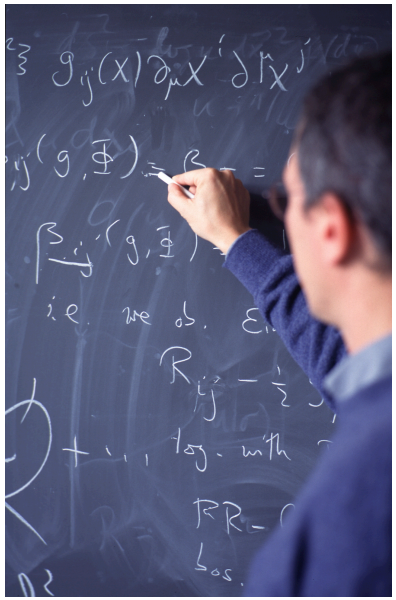
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

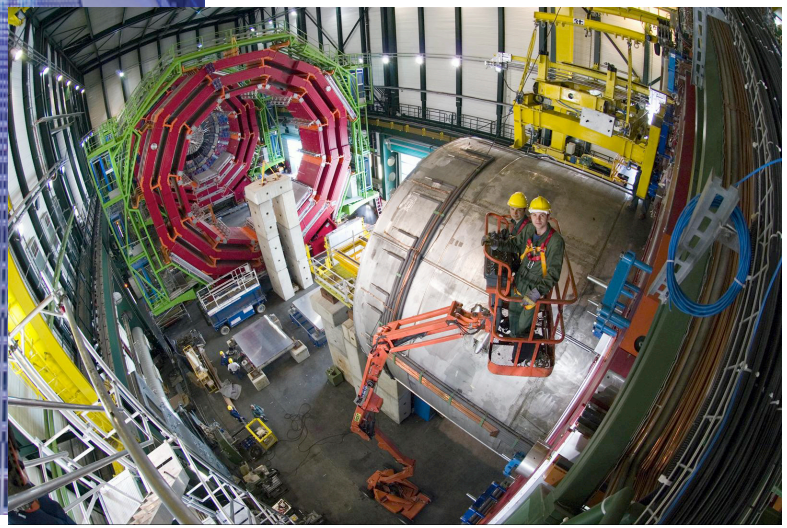


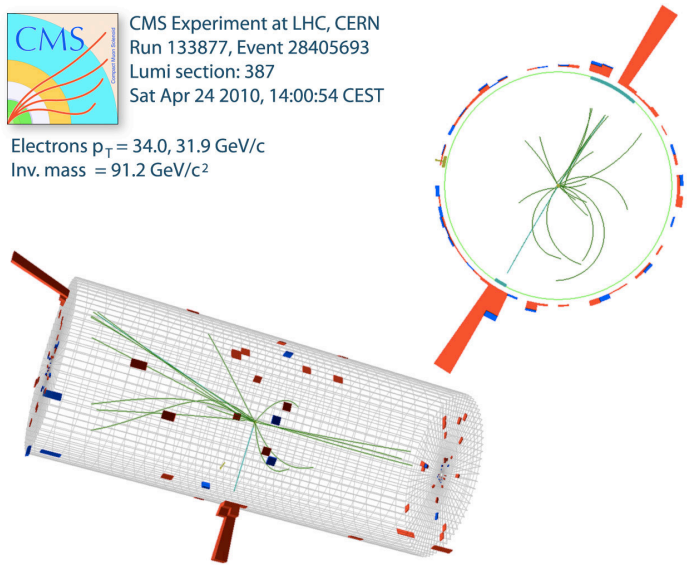
What are all these computers for?

CERN IT Department



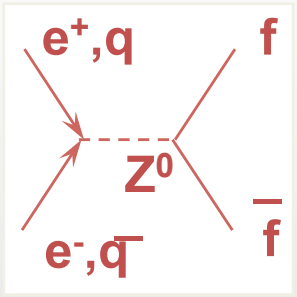


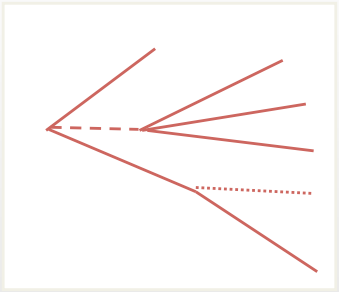


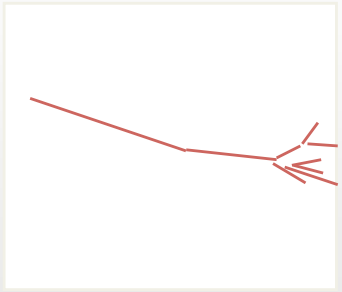


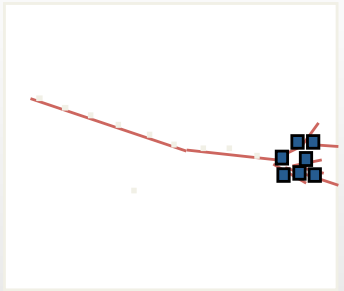
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

From Physics to Raw Data









2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142

BASIC PHYSICS

FRAGMENTATION, DECAY

INTERACTION WITH DETECTOR MATERIAL MULTIPLE SCATTERING, INTERACTIONS

DETECTOR RESPONSE NOISE, PILE-UP, CROSS-TALK, INEFFICIENCY, AMBIGUITY, RESOLUTION, RESPONSE FUNCTION, ALIGNMENT, TEMPERATURE

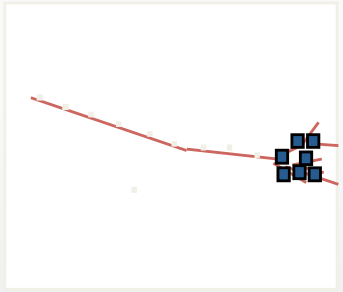
RAW DATA (BYTES)

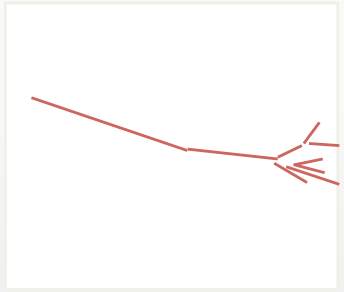
READ-OUT ADDRESSES, ADC, TDC VALUES, BIT PATTERNS

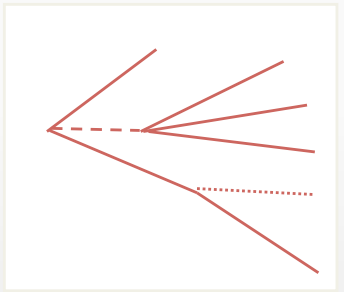
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

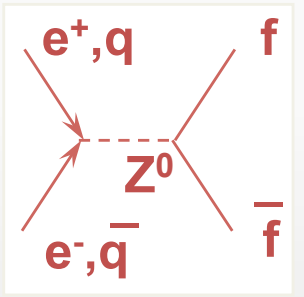
From Raw Data to Physics

CERN IT Department









RAW DATA

CONVERT TO PHYSICS QUANTITIES

DETECTOR RESPONSE
APPLY CALIBRATION, ALIGNMENT,

INTERACTION WITH DETECTOR MATERIAL
PATTERN, RECOGNITION, PARTICLE IDENTIFICATION

FRAGMENTATION, DECAY
PHYSICS ANALYSIS

BASIC PHYSICS

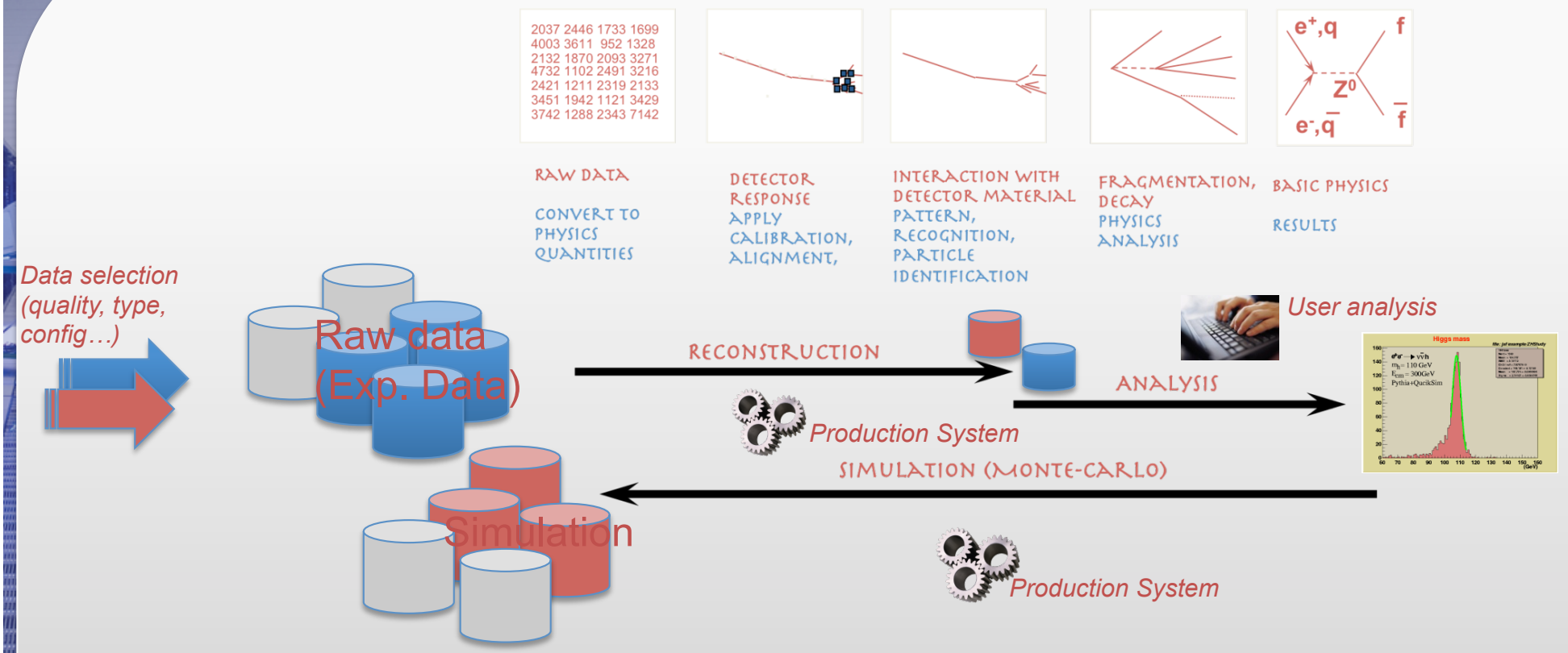
RESULTS

RECONSTRUCTION

ANALYSIS

SIMULATION (MONTE-CARLO)

Analysis flow (user view)

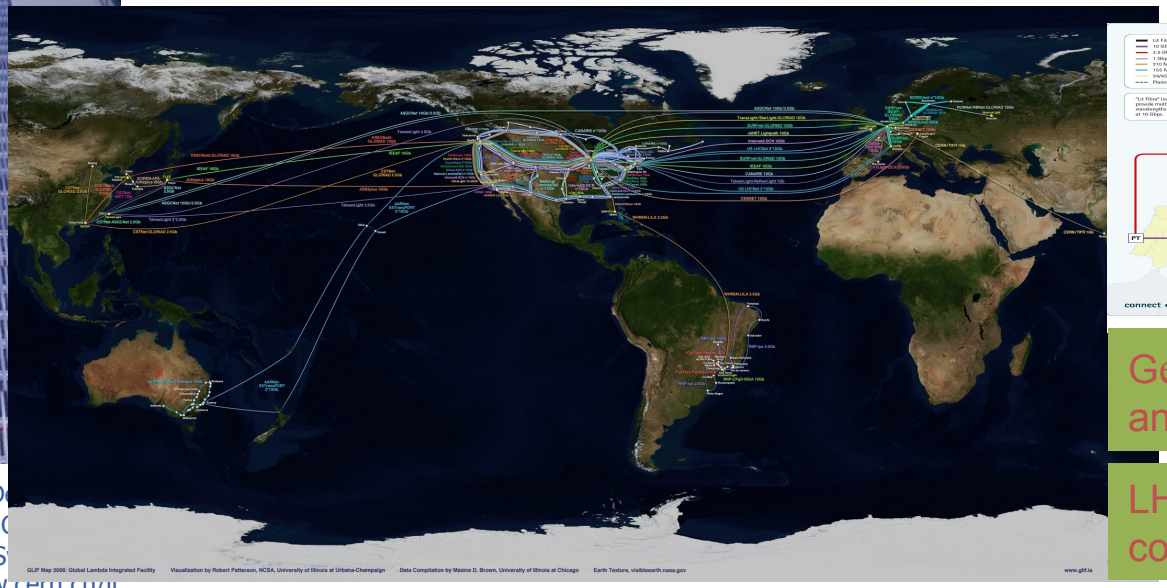


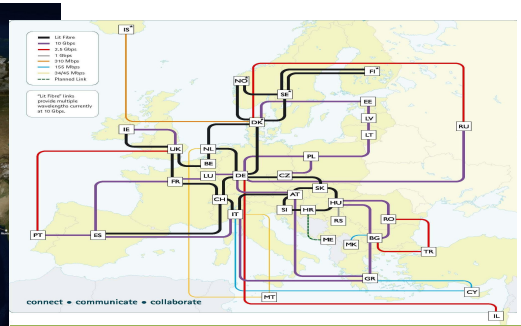
But how this is done *in practice*? Of course we need CPUs, disks, networks etc..
We cannot rush to the solution yet…

Dataset concept = collection of files. Only a small fraction of data in real DBs (e.g calibrations)

CERN Computer Centre: Storage, Distribution and Processing (Reconstruction and Analysis)

CERN IT Department





Géant: the pan-european Research and Education Network

LHCOPN: dedicated links with major computer centres worldwide

Service Level Status overview

Home | Search | KPIs | Tags | Admin | Documentation | Help

Computer Centre By Numbers

23 Mar 2011 Wed 21:30:26

Service information

full name: **Computer Centre By Numbers**

short name: CCBYNUM

group: IT-CF-FPP

site: CERN

email: **imre.szebenyi@cern.ch**

manager: **Imre Szebenyi**

Service availability (more)

availability:

percentage: 100%

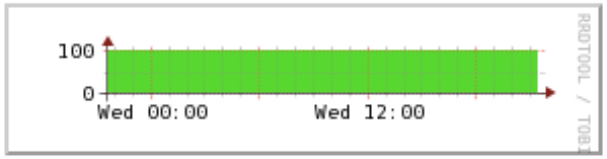
status: **available**

last update: 21:20:34, 23 Mar 2011 (10 minutes ago)

expires after: 1440 minutes

rss feed with status changes

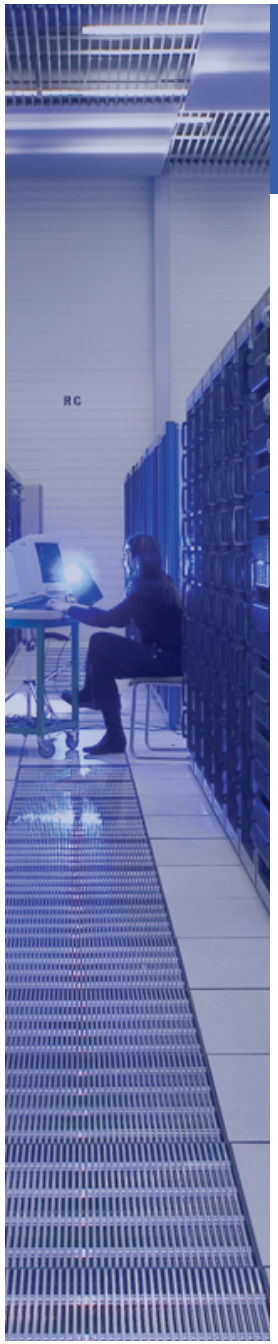
availability in the last 24 hours (more):



Additional service information (more)

Number of processors:	15,678
Number of cores:	64,198
Memory capacity (TiB):	158
Memory modules:	55,990
Raw HDD capacity (TiB):	63,254
Number of HDD's:	64,074
Number of systems:	11,730
Number of RAID controllers:	3,742
Number of enclosures:	1,417
SPEC CPU2006:	482,431
Number of racks:	827
Number of virtual machines:	1,624
Number of Fibre channel ports:	828
Number of 1G ports:	16,936
Number of 10G ports:	558
Current power consumption (kW):	2,506
Current power consumption (kVA):	2,640

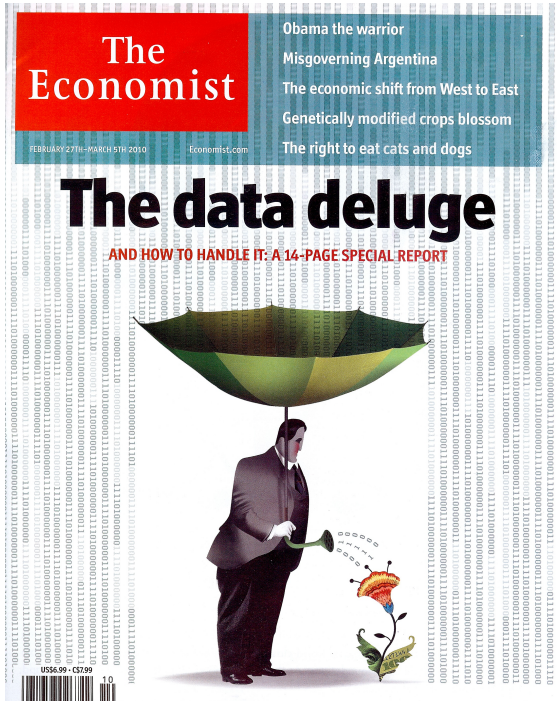
- 24x7 operator and system admin support
- Management and Automation framework for large scale Linux clusters
- Hardware installation & retirement
  - ~7,000 hardware movements/year; ~1000 disk failures/year



The LHC Data Challenge

CERN IT Department

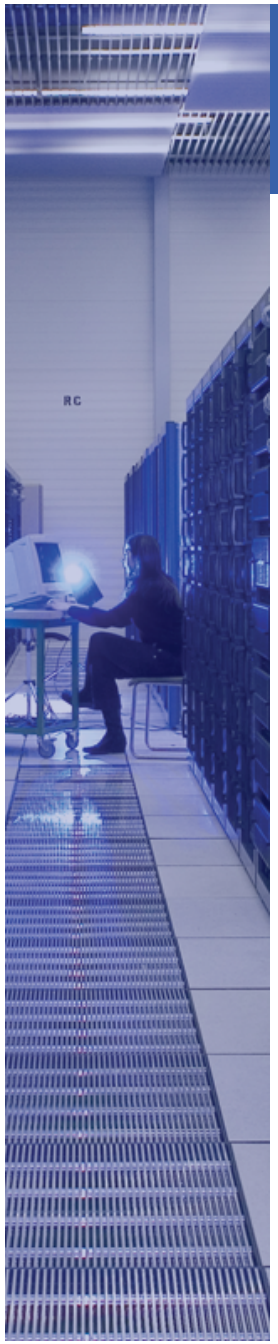
- The accelerator will run for 20 years
- Experiments ***are*** producing about **15 Million Gigabytes** of data each year (about 20 million CDs!)
- LHC data analysis requires a computing power equivalent to **~100,000 of today's fastest PC processors**
- Requires many cooperating computer centres, as CERN can *only* provide ~20% of the capacity



A challenge for physics…

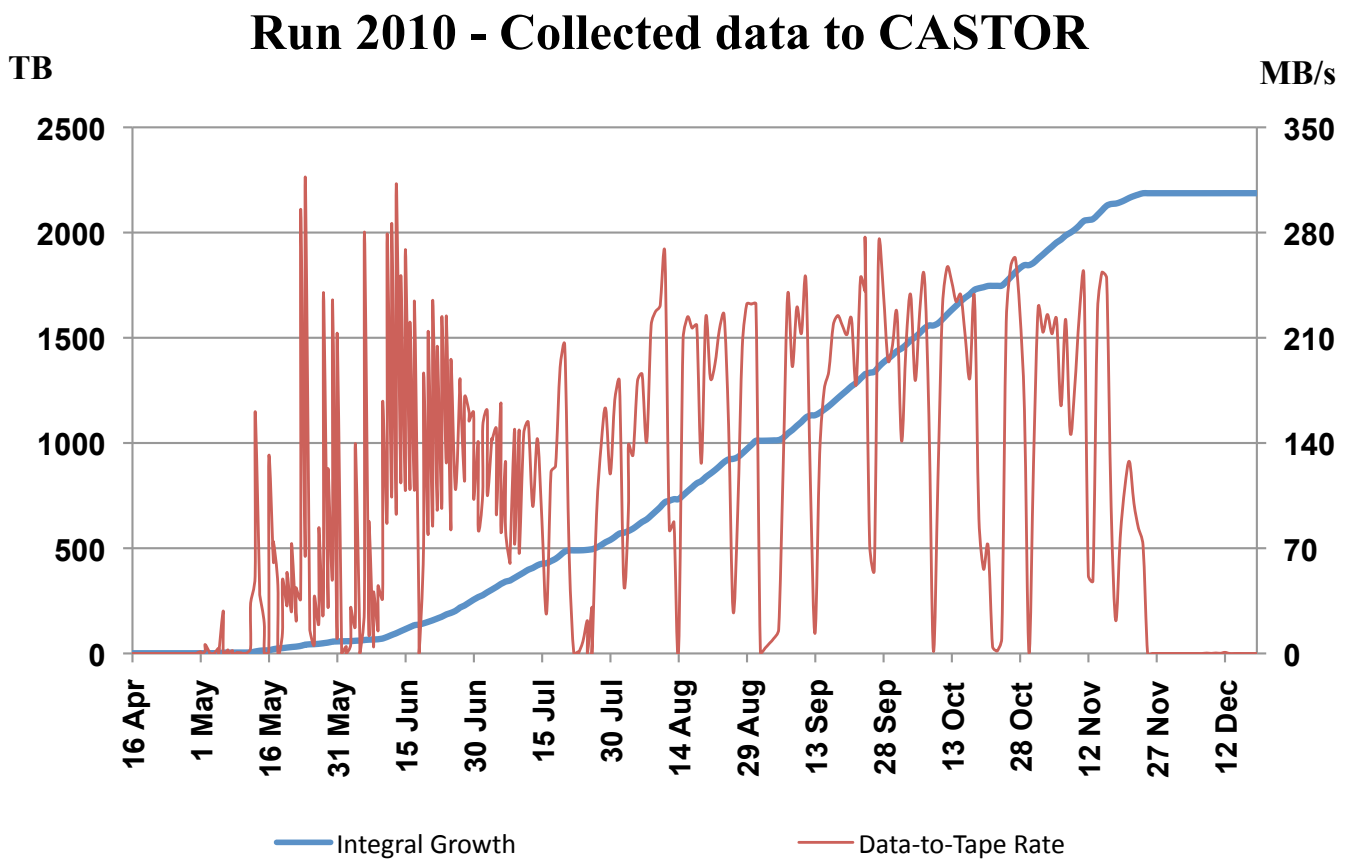
… and a challenge for technology research and industry as well

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

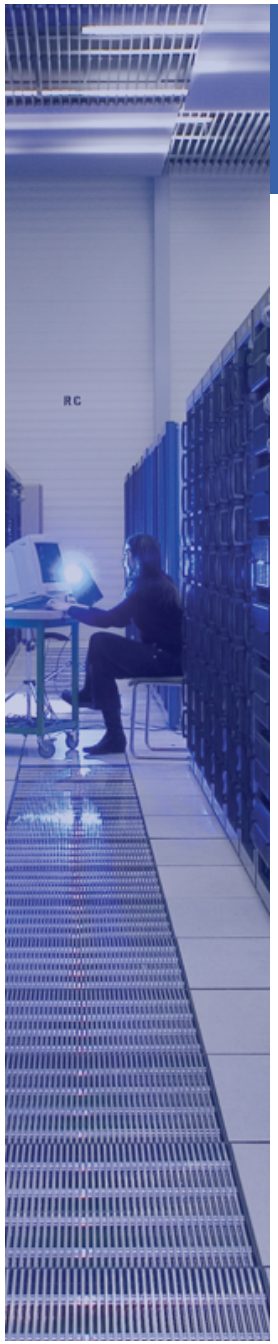


COMPASS in 2011

CERN IT Department



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

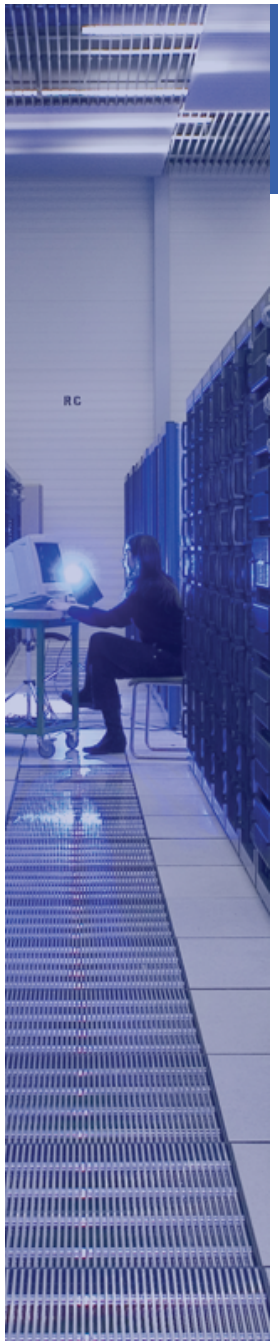


“Historical” example…

CERN IT Department

1990s:
The web was invented at CERN!
The machine used by Tim Berners-Lee in 1990 to develop and run the first WWW server, multi-media browser and web editor.

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it



A new solution: the Grid

CERN IT Department

- Use the Grid to unite computing resources of particle physics institutes around the world

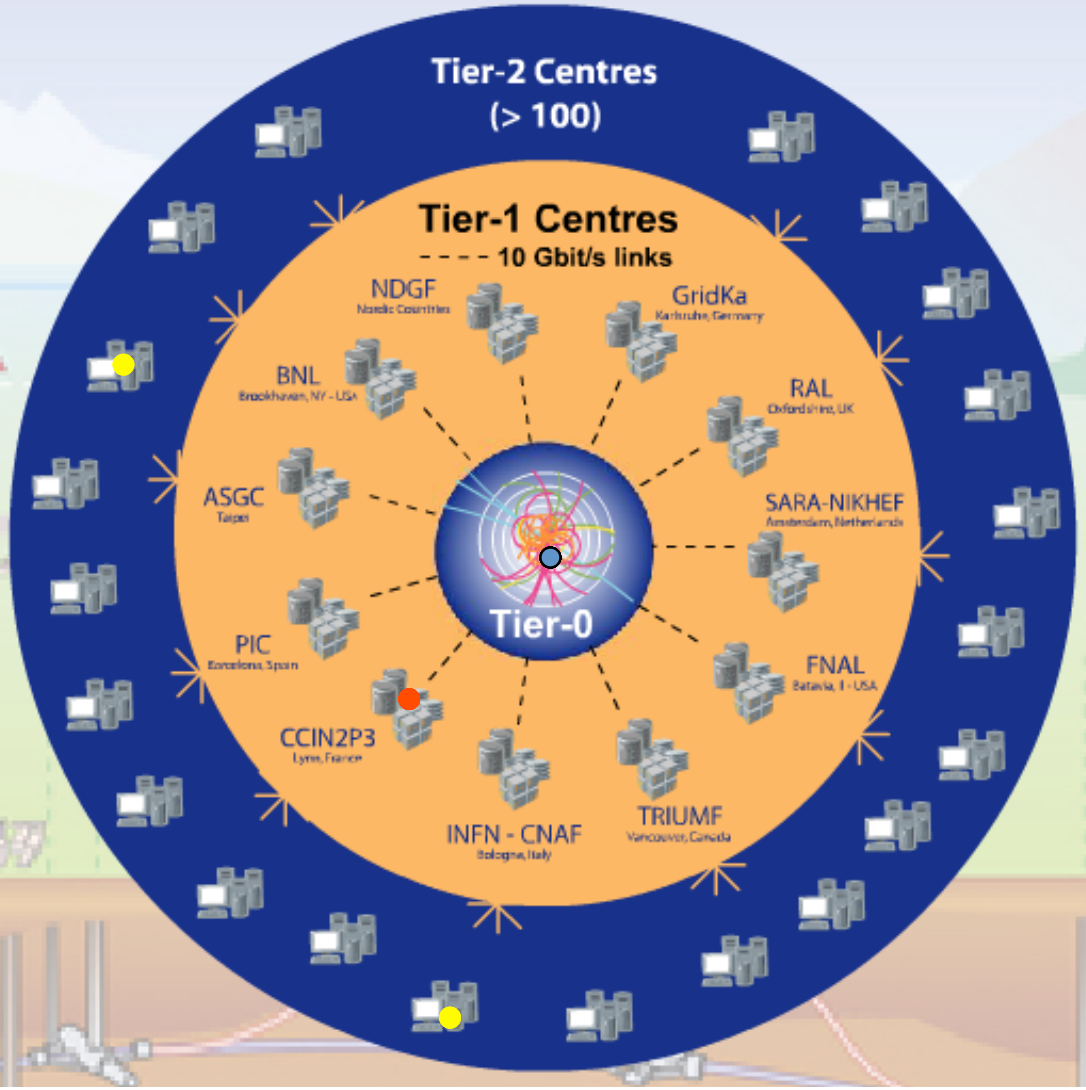
The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations

The **Grid** is an infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

WLCG Tiers Organization

CERN IT Department

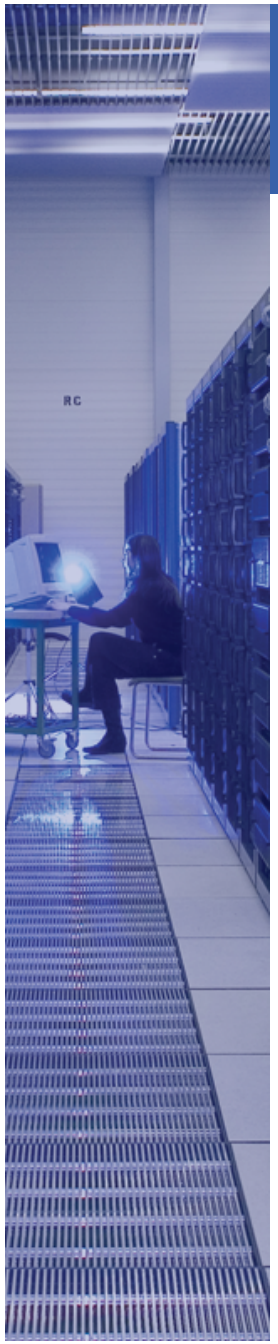


Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (11 centres):
- Permanent storage
- Re-processing
- Analysis

Tier-2 (~130 centres):
- Simulation
- End-user analysis

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

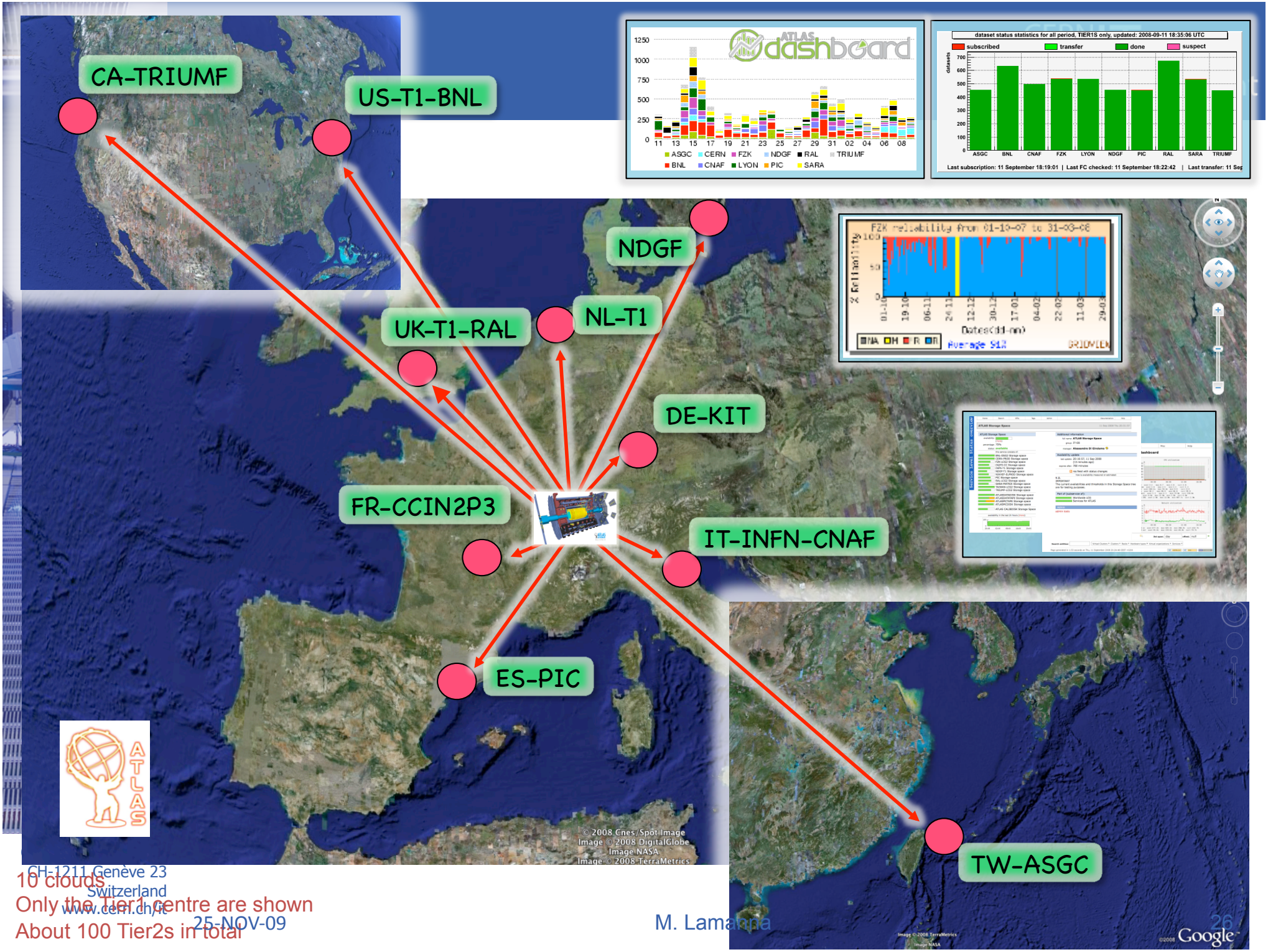


How does it work?

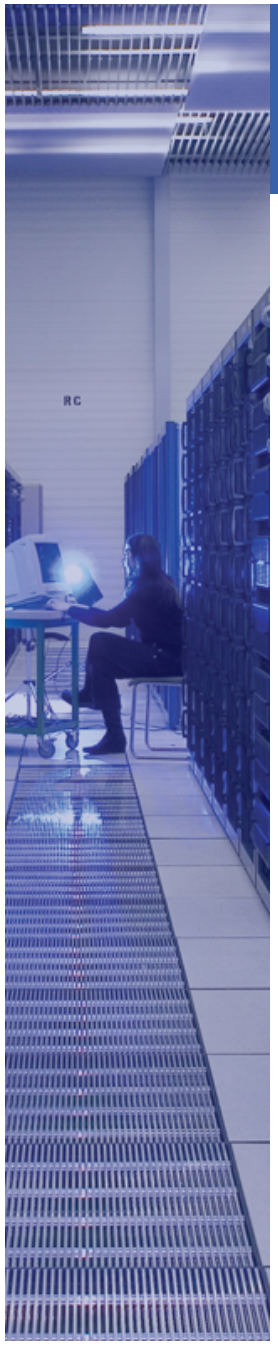
CERN IT Department

- ATLAS
  - Not substantially different for the other HEP experiments
  - Heavily simplified…
- What do we want to achieve
  - The user wants to specify a subset of the data and run applications on it (chain of programs reading intermediate outputs)
  - Only at the end of the chain data sizes and computational complexity this can be (possibly) done on a laptop
    - 1000+ of physicists worldwide after the same data

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

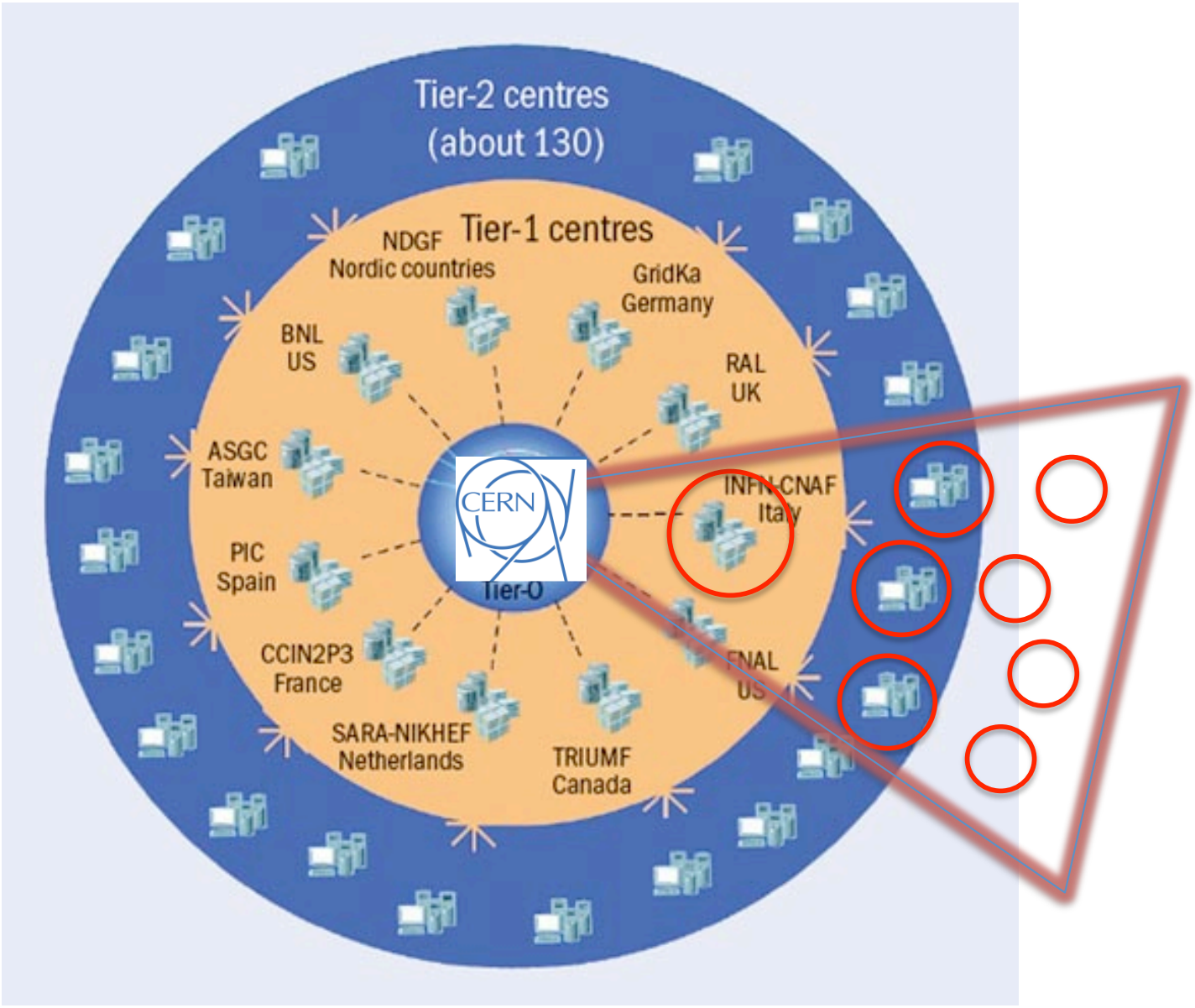


10 clouds
Only the Tier1 centre are shown
About 100 Tier2s in total



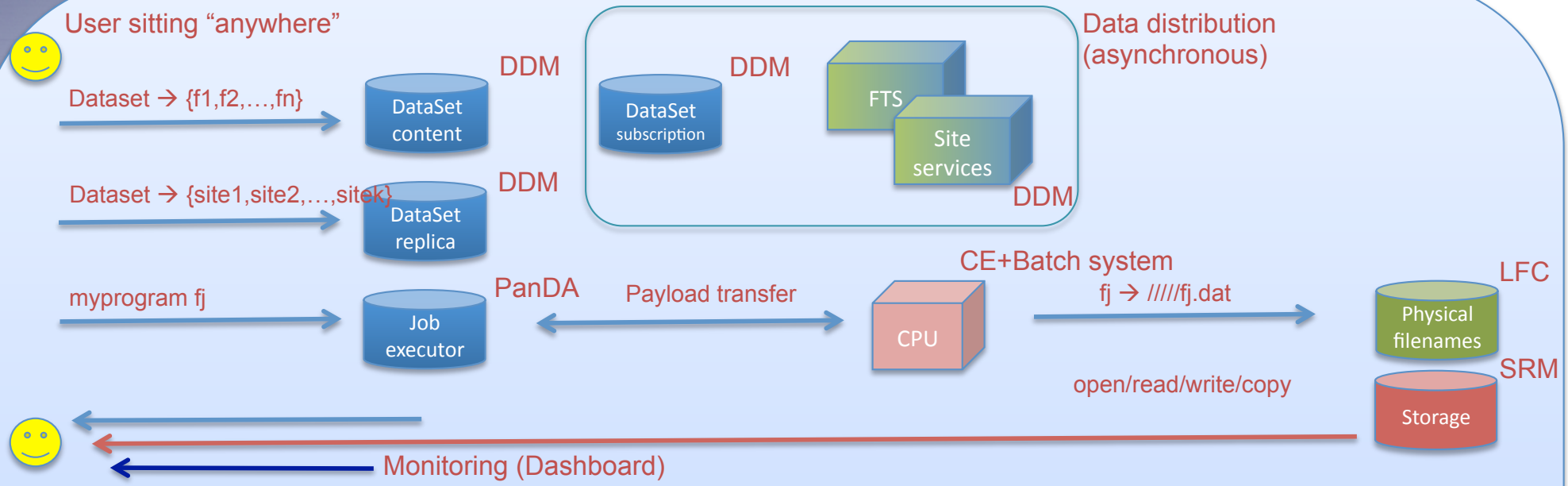
Distributed analysis

CERN IT Department



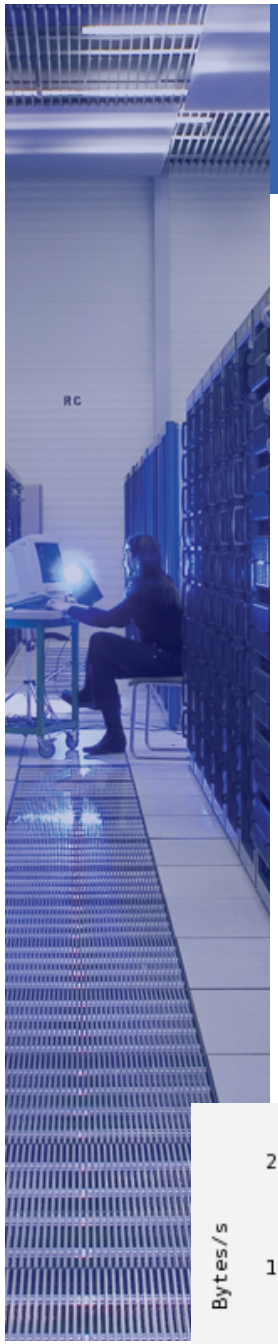
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

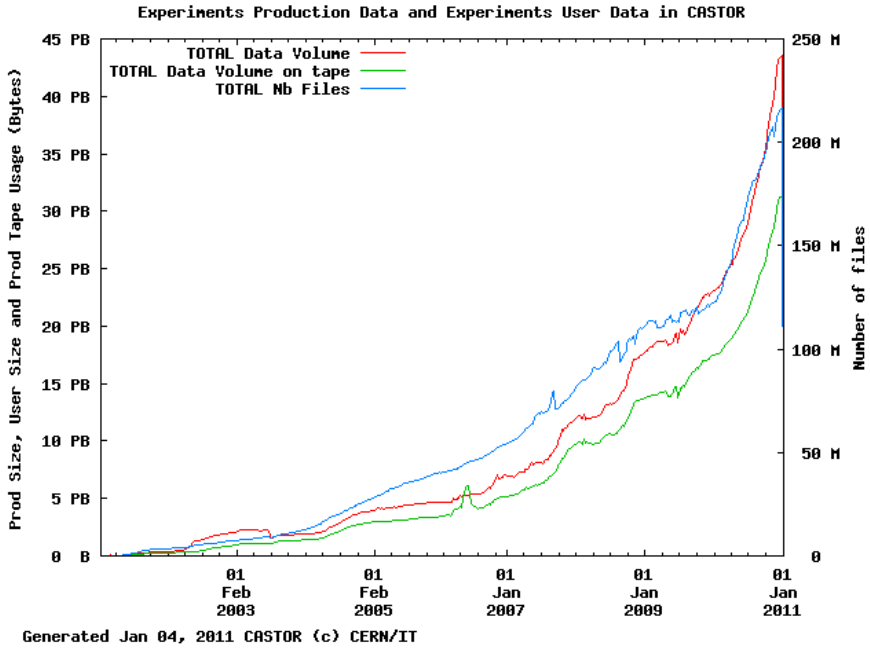
Behind the scenes…



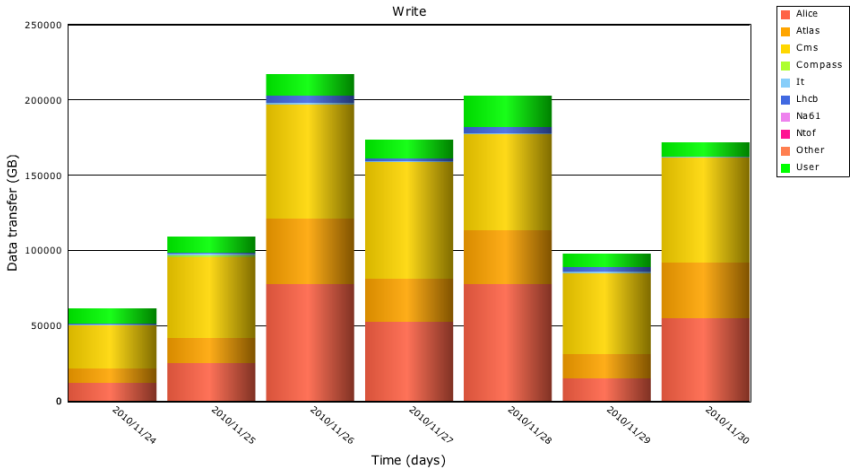
Simplified
One job executed in one cloud

2010 data taking as seen in the CERN CC

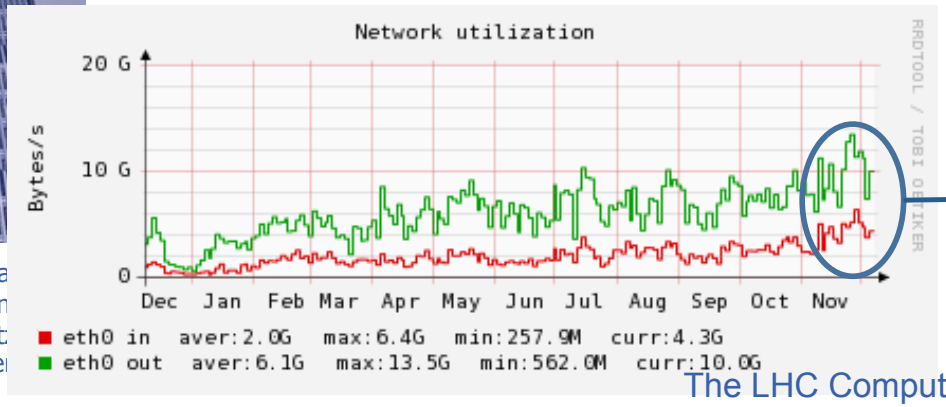


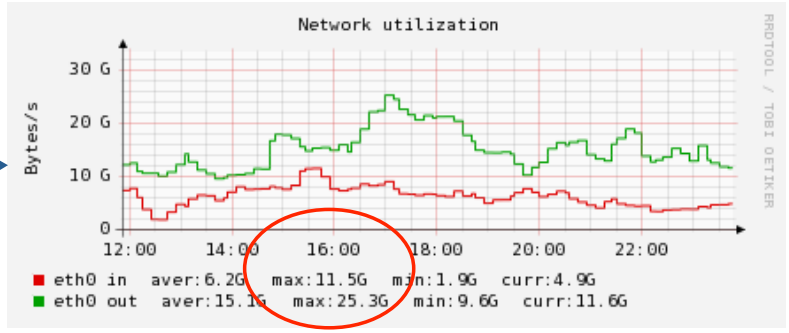


Stored ~ 15 PB in 2010 with peaks at 220 TB/day during Pb +Pb



Tier-0 Bandwidth
Average in: 2 GB/s with peaks at 11.5 GB/s
Average out: 6 GB/s with peaks at 25 GB/s

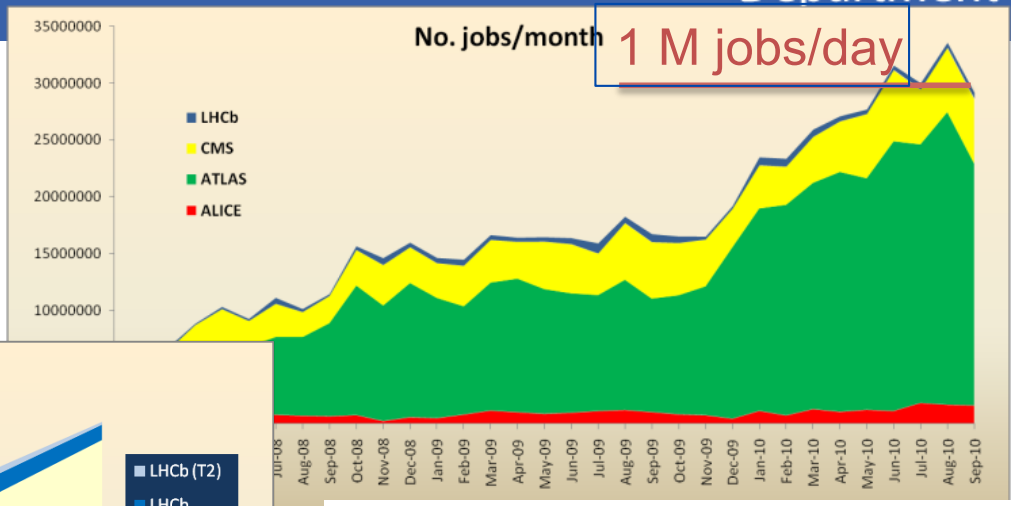


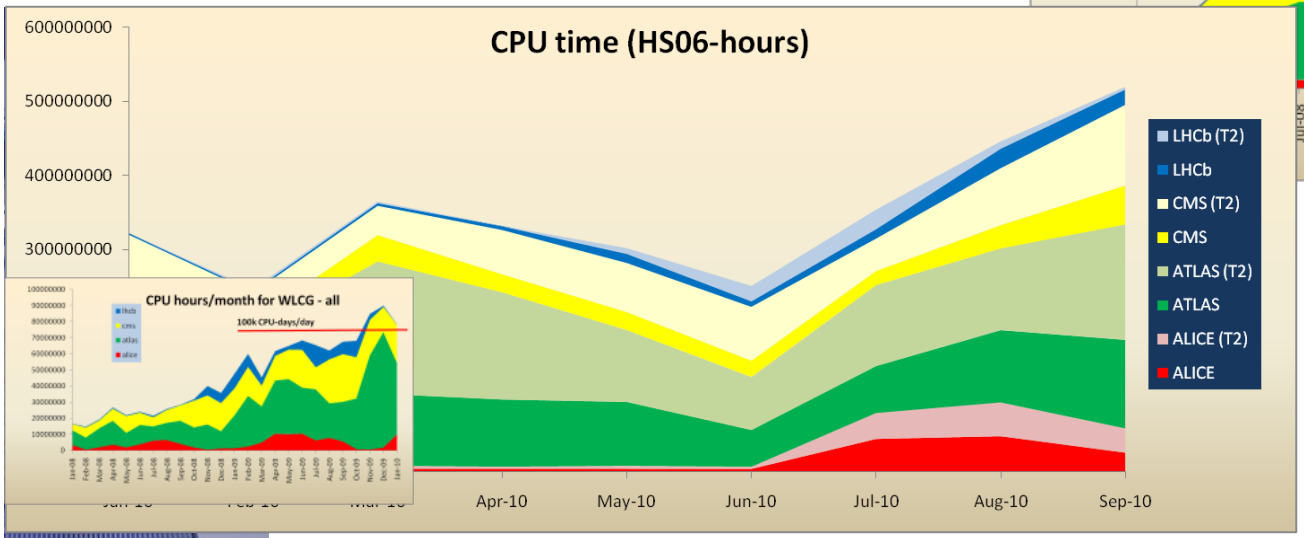


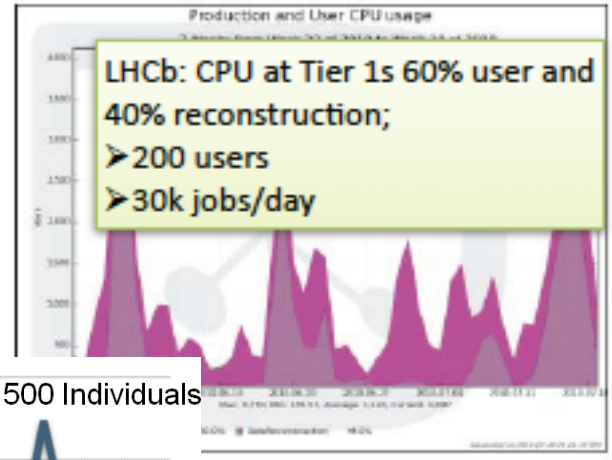
WLCG Usage

CERN IT Department

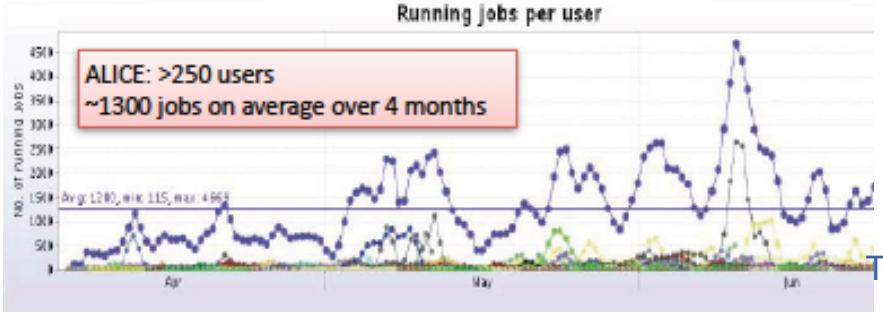
- Use remains consistently high
  - 1 M jobs/day; >>100k CPU-days/day
  - Actually much more inside pilot jobs

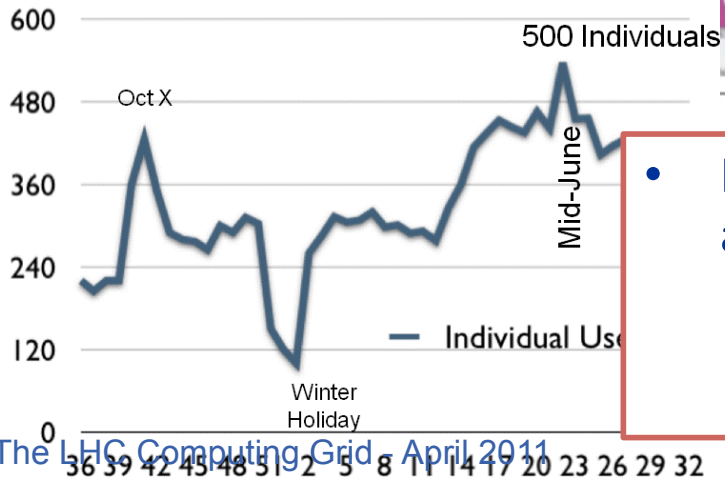




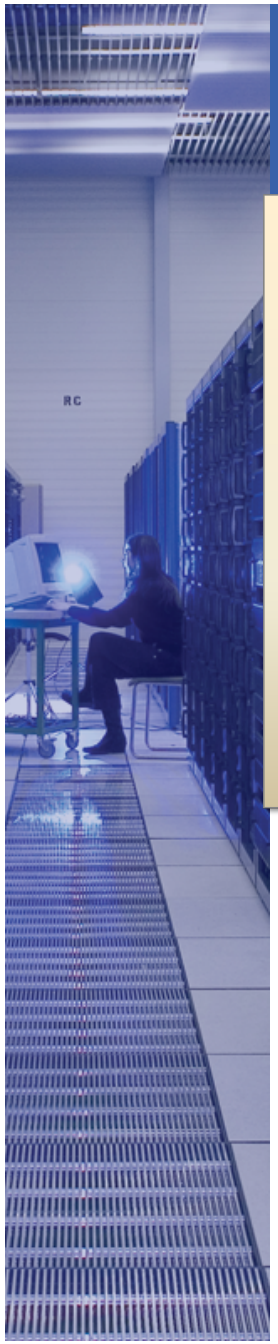


As well as LHC data, large simulation productions ongoing



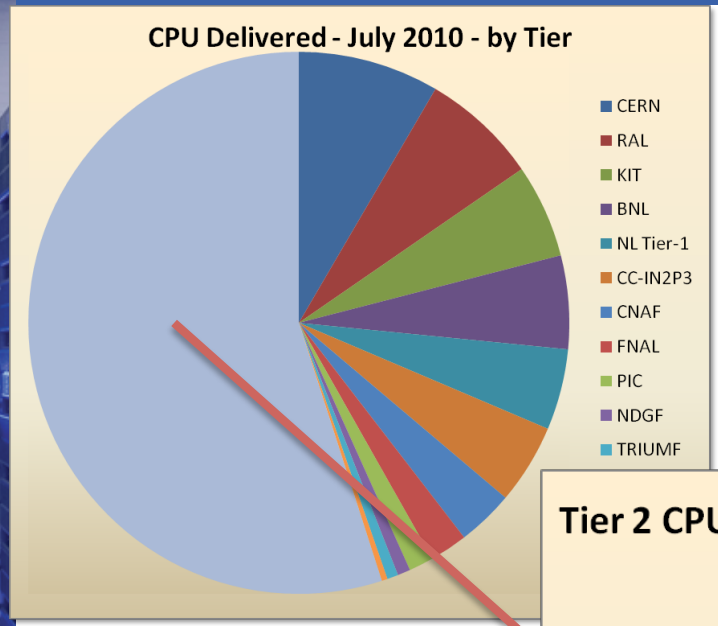


- Large numbers of analysis users
  - CMS ~800,
  - ATLAS ~1000,
  - LHCb/ALICE ~250



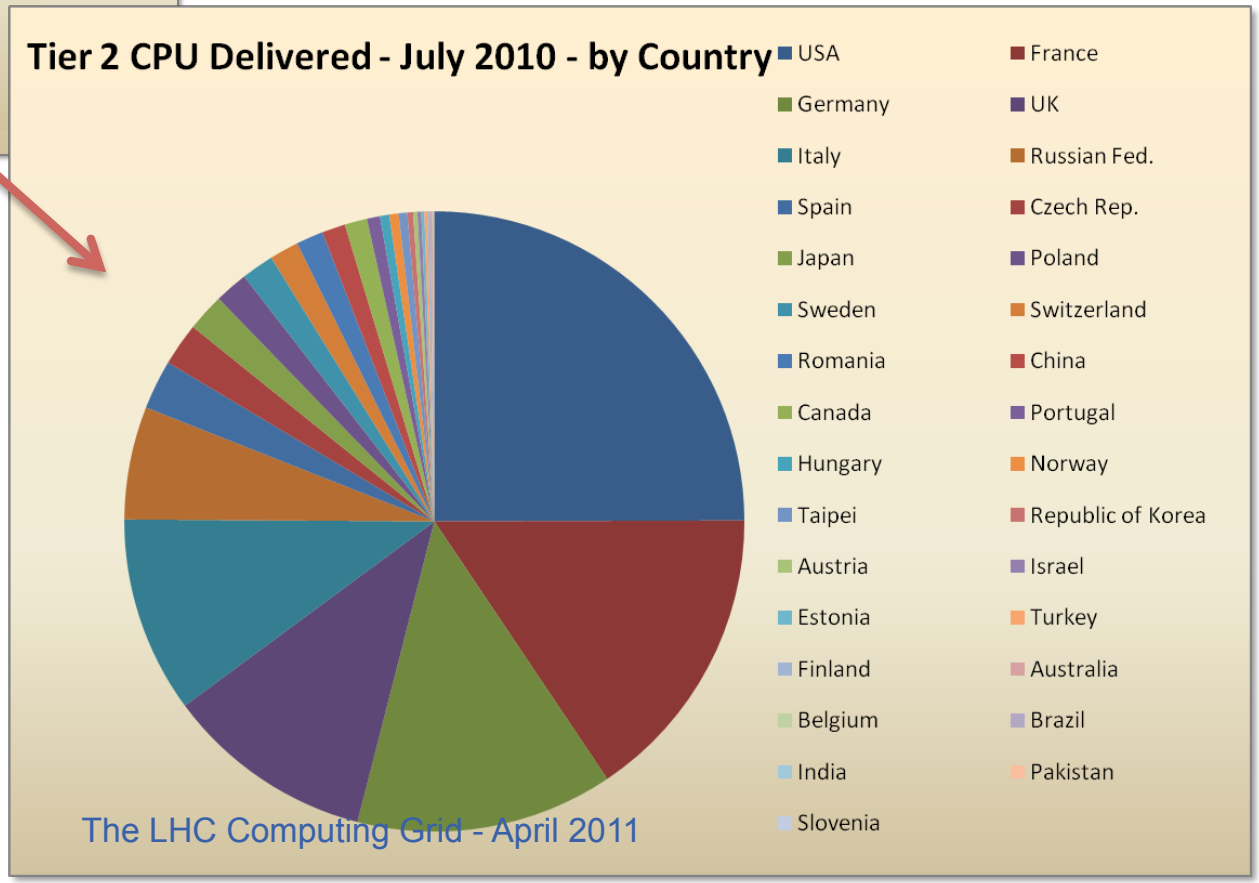
Cooperating effort

CERN IT Department



- Significant use of Tier 2s for analysis
  - Frequently-expressed concern that too much analysis would be done at CERN is not reflected

- Tier 0 capacity underused in general
  - But this is expected to change as luminosity increases



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Impact of the LHC Computing Grid in Europe

CERN IT Department

eGee Enabling Grids for E-sciencE

Scheduled = 21539
Running = 25374

- LCG has been the driving force for the European multi-science Grid EGEE (Enabling Grids for E-sciencE)
- EGEE was a global effort, and the largest Grid infrastructure worldwide
- Co-funded by the European Commission (Cost: ~170 M€ over 6 years, funded by EU ~100M€)
- Now moving to a permanent European Infrastructure: EGI

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
…

>300 sites
48 countries
>200,000 cores
>20 PetaBytes
>10,000 users
>150 VOs
>150,000 jobs/day

21:13:50 UTC

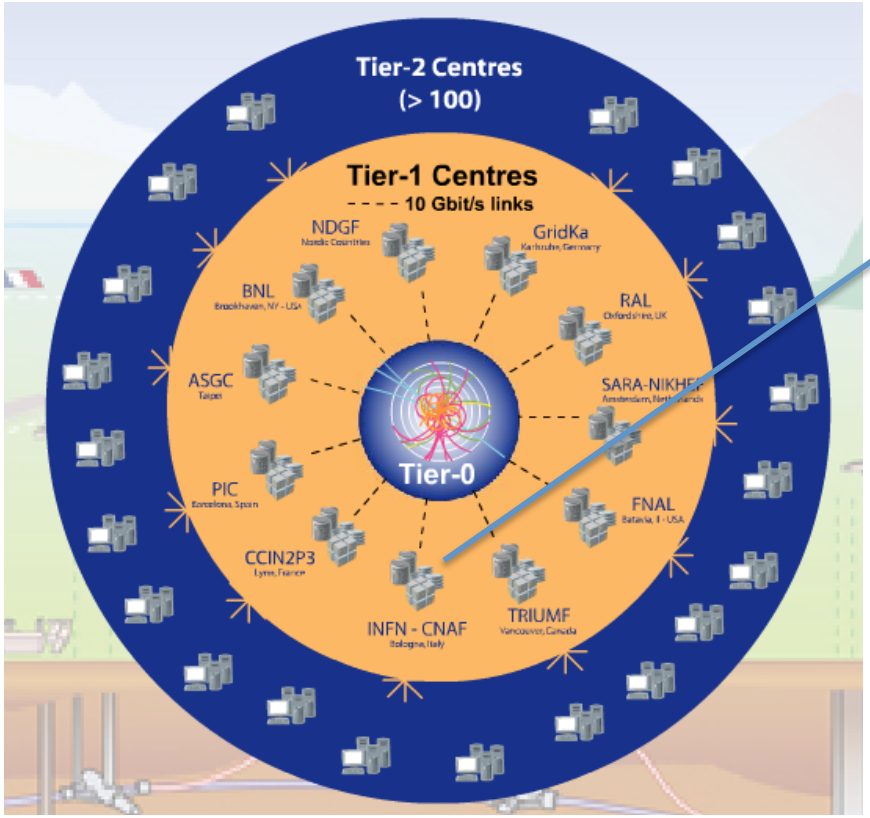
GridPP UK Computing for Particle Physics

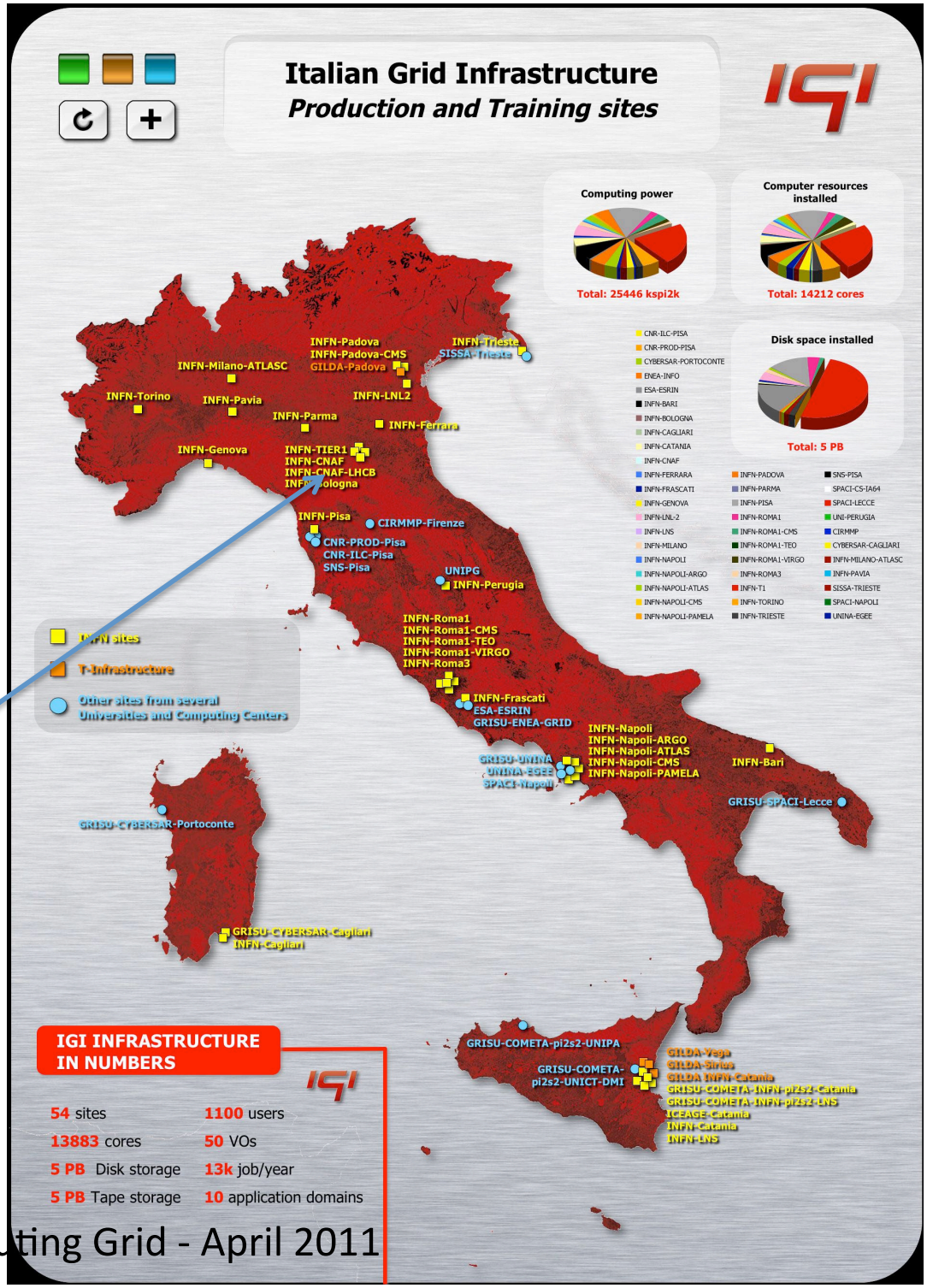
More info:

INFN (Istituto Nazionale Fisica Nucleare):

http://www.infn.it

IGI (Italian Grid Initiative): http://
www.italiangrid.org/





...more communities sharing tools designed for LHC and being useful in other domains!

ATLAS

LHCb

Cambridge Ontology

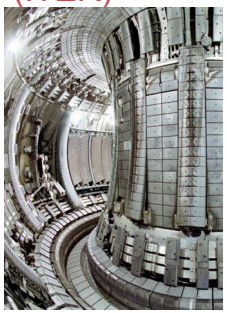
Geant 4

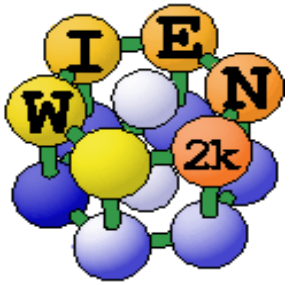
COMPASS

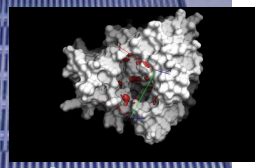
Garfield

International Telecommunication Union

Nuclear fusion (ITER)



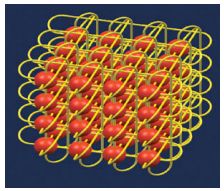




HARP

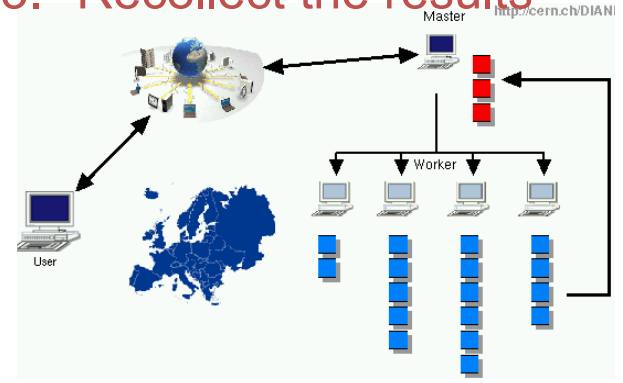
med austron

Lattice QCD



Always reuse the same pattern:

1. Divide the problem in small task that can be distributed independently
2. Distribute these tasks to “dumb” workers (Grid processes)
3. Recollect the results



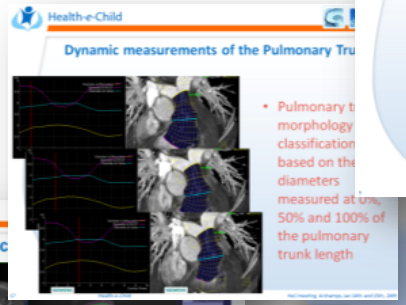
Academia Sinica Genomics Research Center
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

http://cern.ch/ganga

http://cern.ch/DIANE

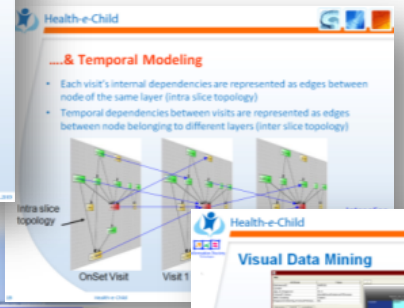
Information Society Technologies

Measurement of Pulmonary Trunk



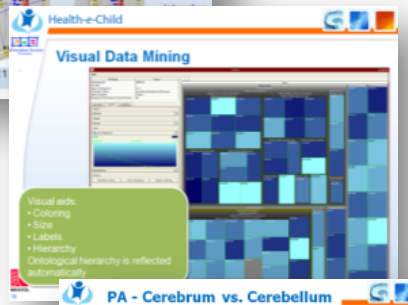
Similarity Search

Temporal Modelling



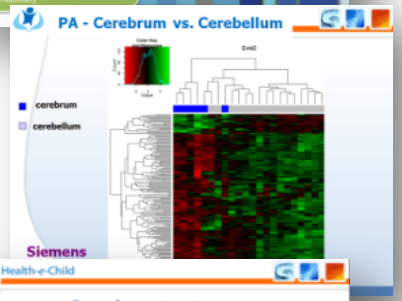
RV and LV Automatic Modelling

Visual Data Mining

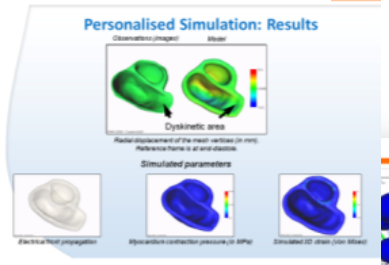


Surgery Planning

Genetics Profiling

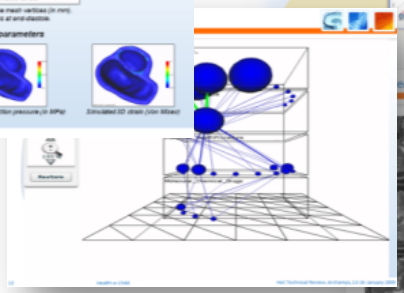


Personalised Simulation

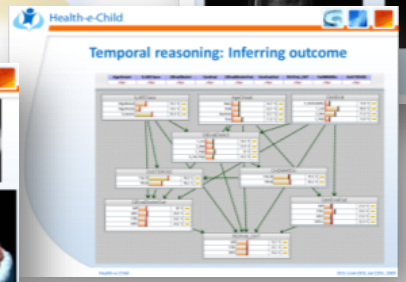


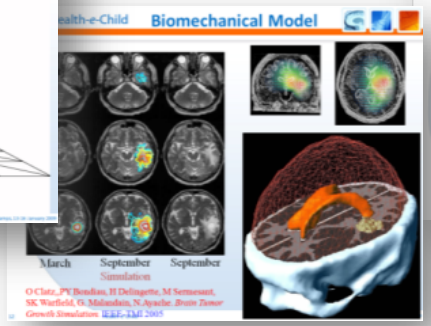
Treatment Response

Semantic Browsing



Inferring Outcome





Biomechanical Models
Tumor Growth Modelling

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

ITU conference (2006)

The problem:
Assign frequencies for digital radio and television (international treaty)

Critical point:
Need on dependability: verify (iteratively) the compatibility between radio stations

Solution:
Use the EGEE grid + a system used in ATLAS and LHCb to increase the reliability of the Grid

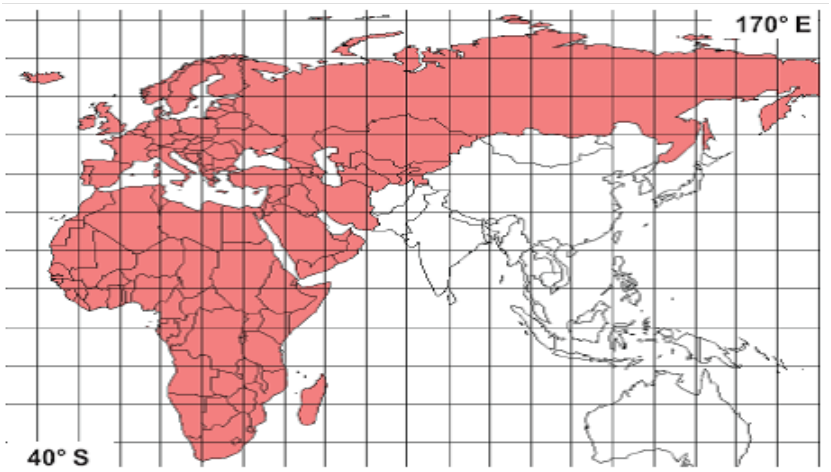
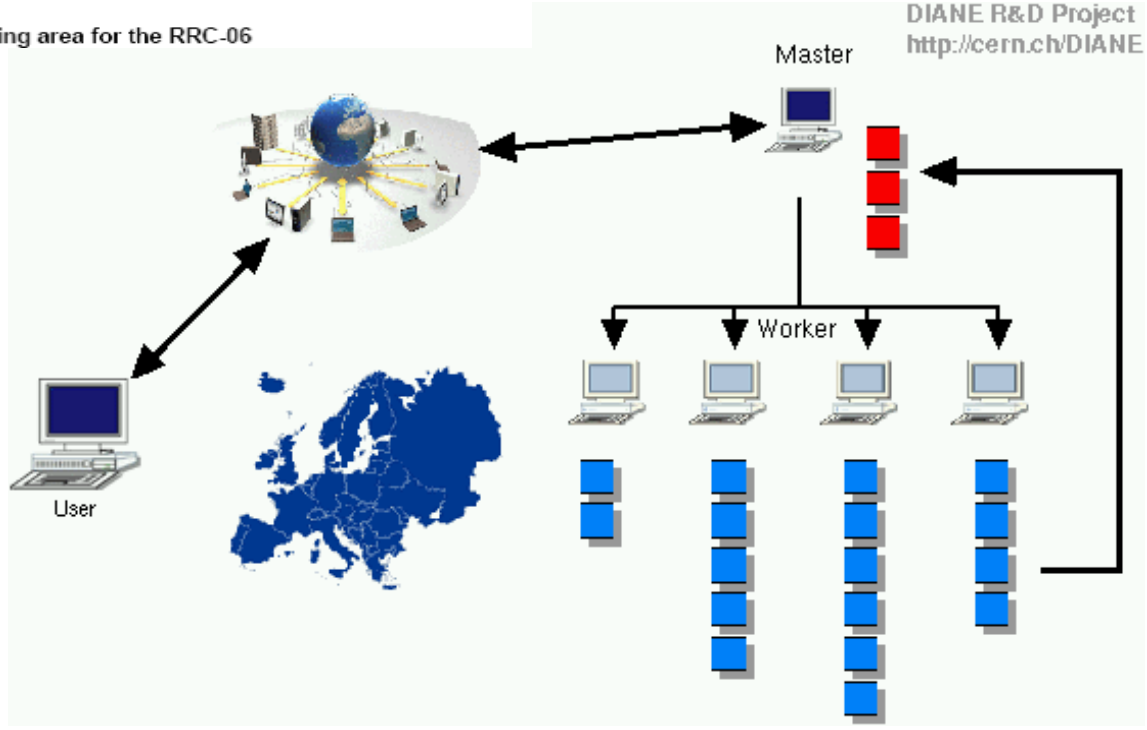


Figure 1
The extent of the planning area for the RRC-06



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Soil & Water Assessment Tool | SWAT

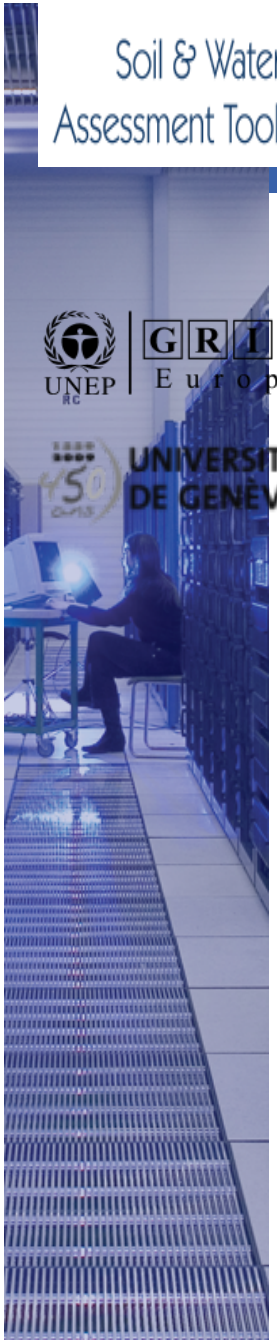
SEVENTH FRAMEWORK PROGRAMME

enviroGRIDS

UNEP | GRID Europe

UNIVERSITÉ DE GENÈVE

Use the grid to study large parameter spaces and simulate different scenarios



Infiltration/plant uptake/ Soil moisture redistribution

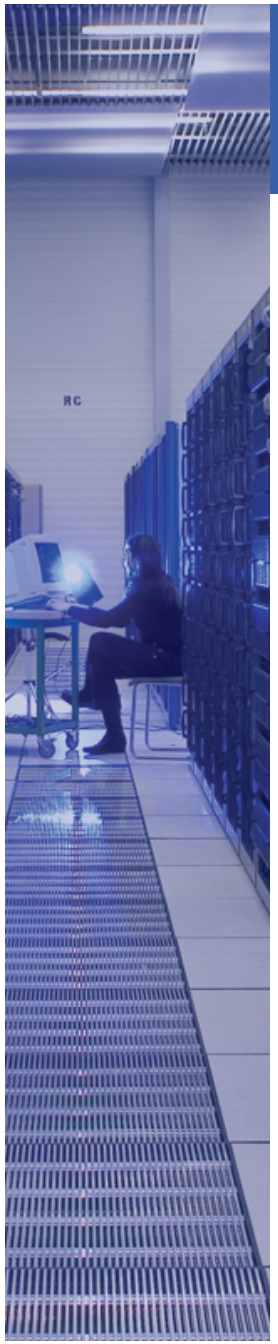
Revap from shallow aquifer

Percolation to shallow aquifer

Flow out of watershed

Recharge to deep aquifer

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it



For more information about the Grid:

www.cern.ch/lcg

www.eu-egee.org

www.eu-egi.org/

www.gridcafe.org

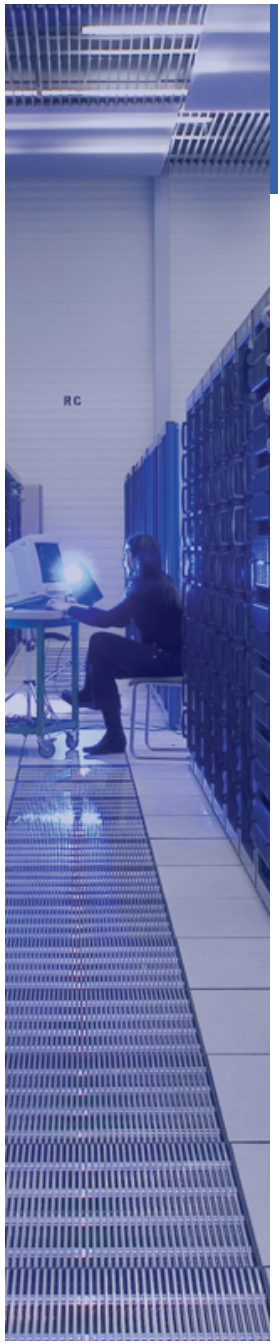
Thank you for your kind attention!

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CERN options for students

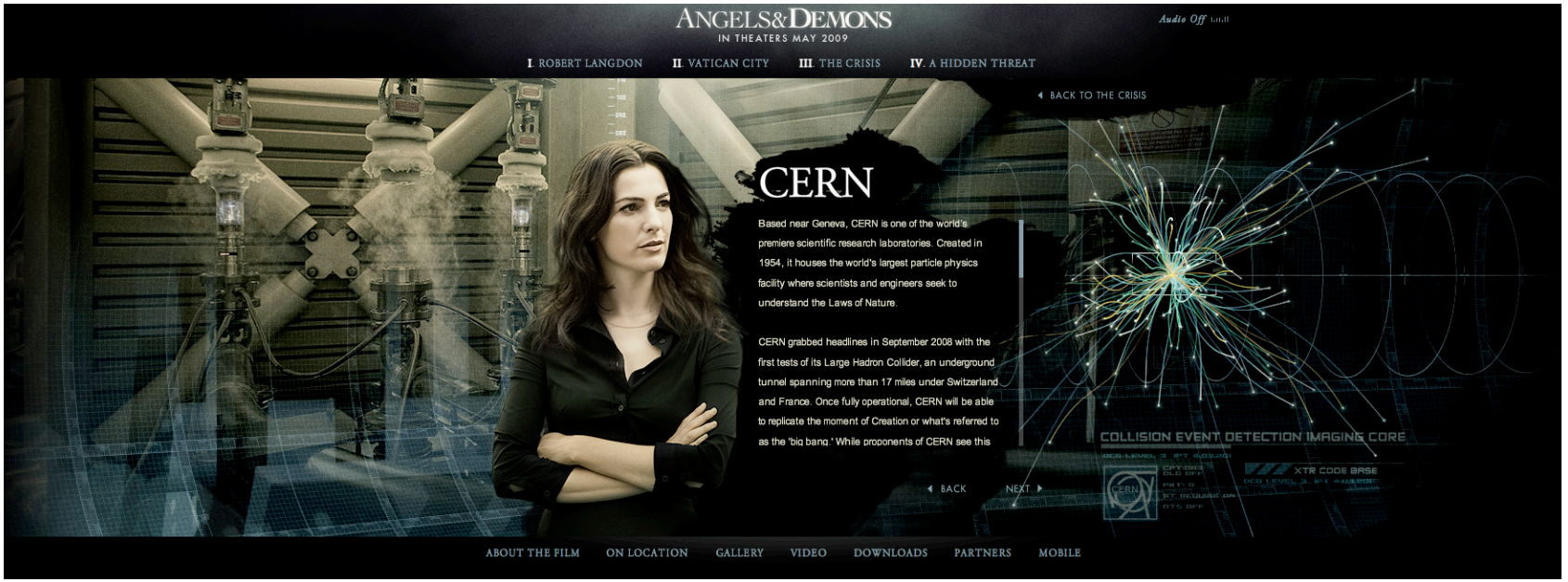
- University level (BS/Master)
  - Summer student
  - OpenLab summer students
- Master thesis
  - Technical student (non physicist)
- PhD students
  - Doctoral students
- Young scientists/engineers
  - Fellowship
  - Other programmes

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it



Questions?

CERN IT Department



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it