
CERN

Introduction to CERN and upcoming challenges in Scientific Computing

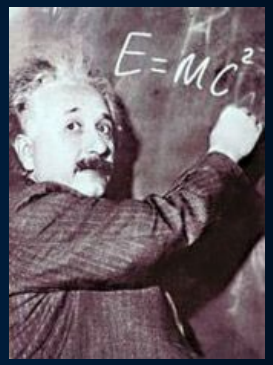
Frédéric Hemmer
CERN - IT Department

Outline

- Overview of CERN
- WLCG, CERN IT & Big data
- Data Preservation

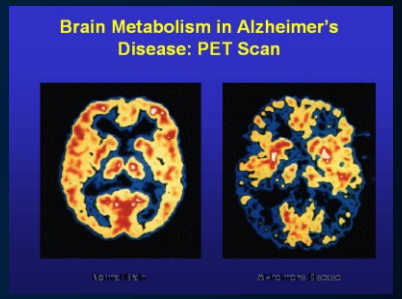
What is CERN?

The Mission of CERN



- **Push back** the frontiers of knowledge

  E.g. the secrets of the Big Bang …what was the matter like within the first moments of the Universe's existence?



- **Develop** new technologies for accelerators and detectors

  Information technology - the Web and the GRID

  Medicine - diagnosis and therapy

- **Train** scientists and engineers of tomorrow

- **Unite** people from different countries and cultures

CERN: founded in 1954: 12 European States
"Science for Peace"
Today: 21 Member States

~ 2300 staff
~ 1300 other paid personnel
~ 11500 scientific users

Budget (2015) ~1000 MCHF

Member States: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland and United Kingdom

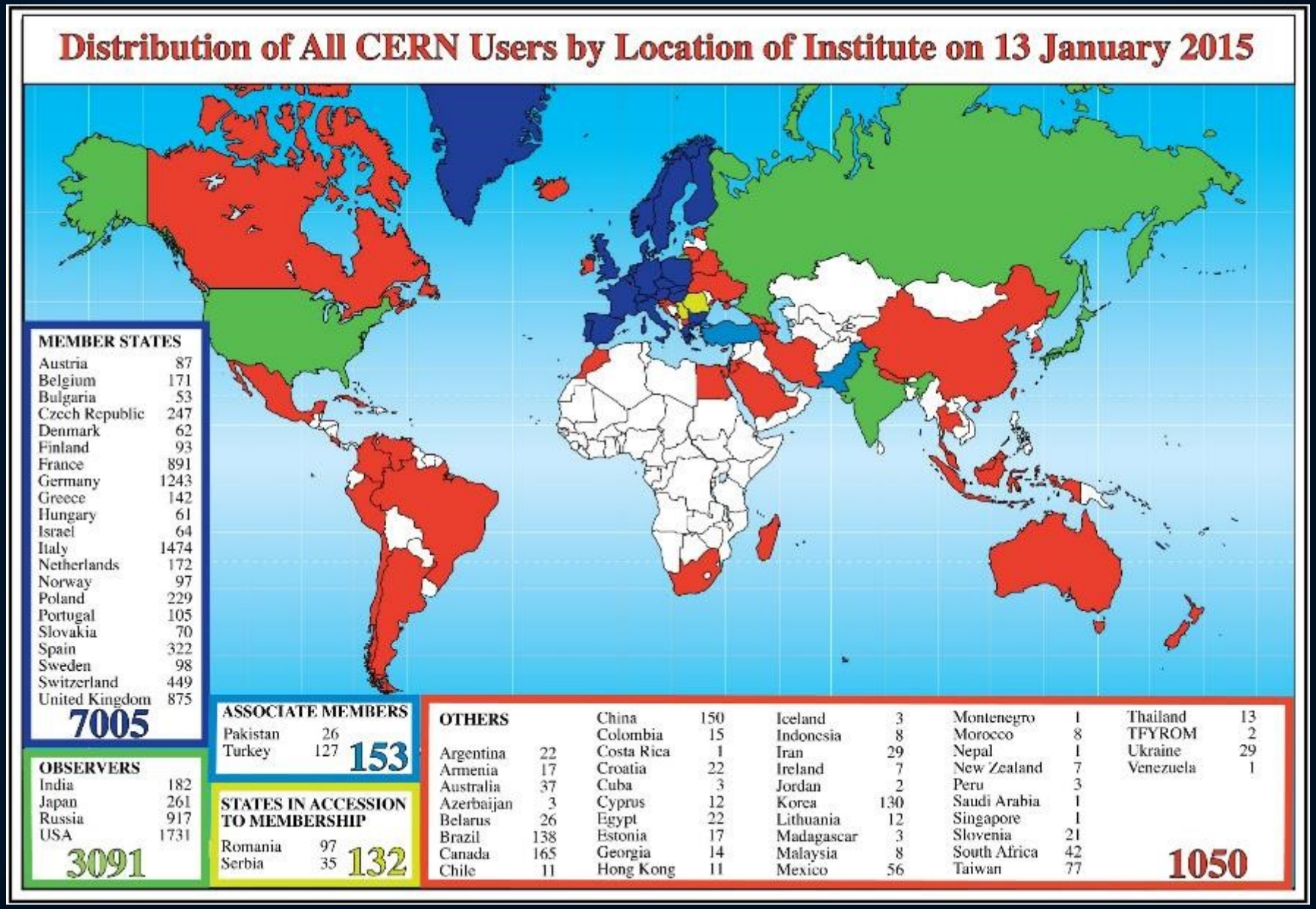
Associate Member States: Pakistan, Turkey

States in accession to Membership: Romania, Serbia

Applications for Membership or Associate Membership: Brazil, Croatia, Cyprus, India, Russia, Slovenia, Ukraine

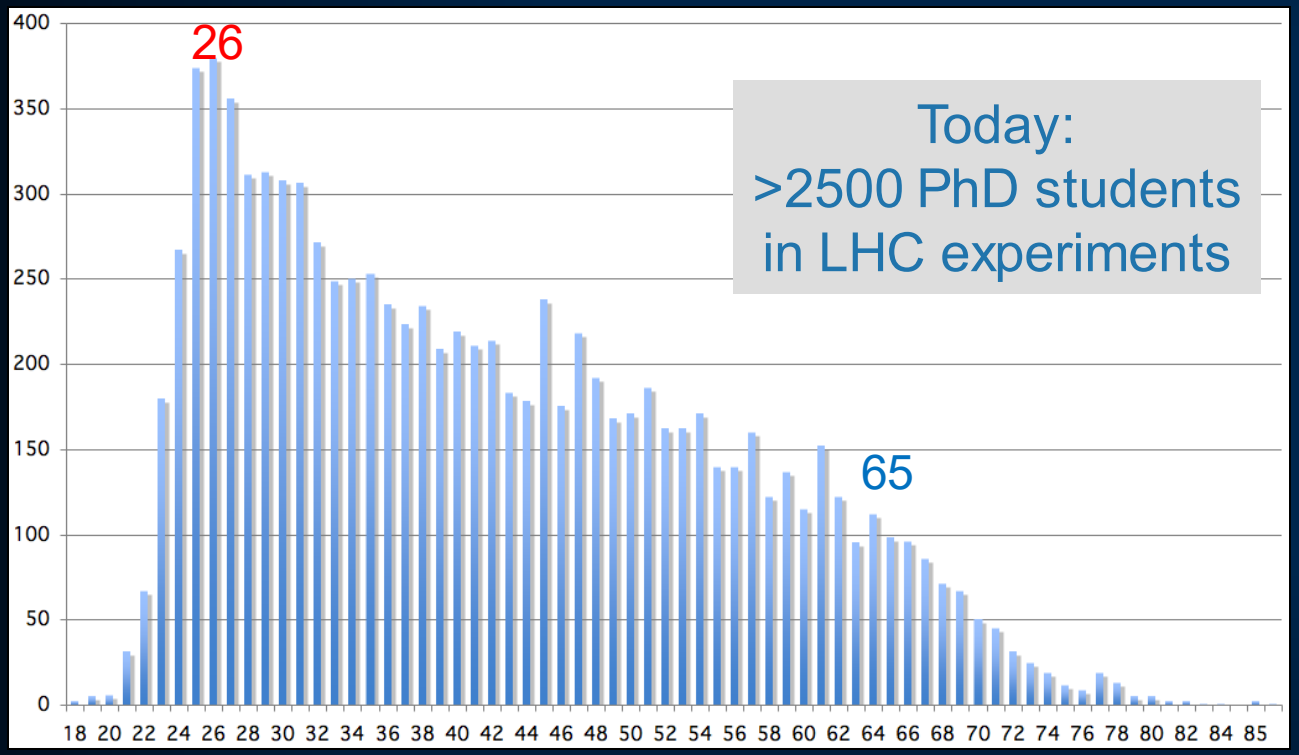
Observers to Council: India, Japan, Russia, United States of America; European Union, JINR and UNESCO

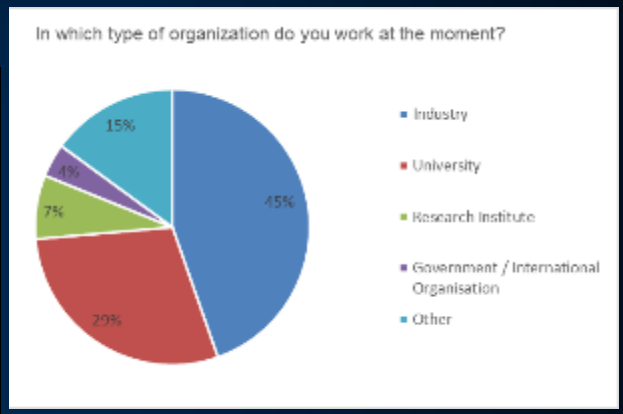
Science is getting more and more global

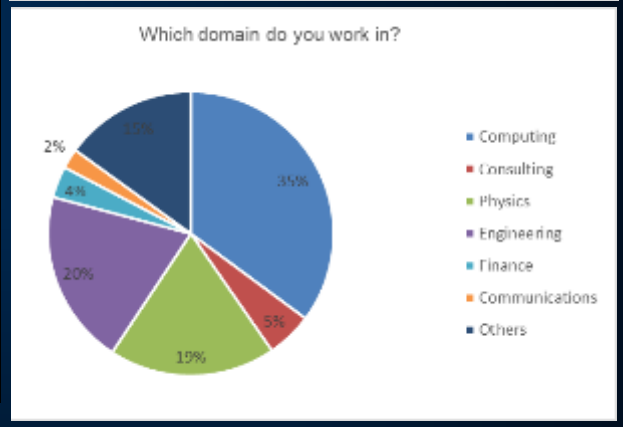


Age Distribution of Scientists

- and where they go afterwards



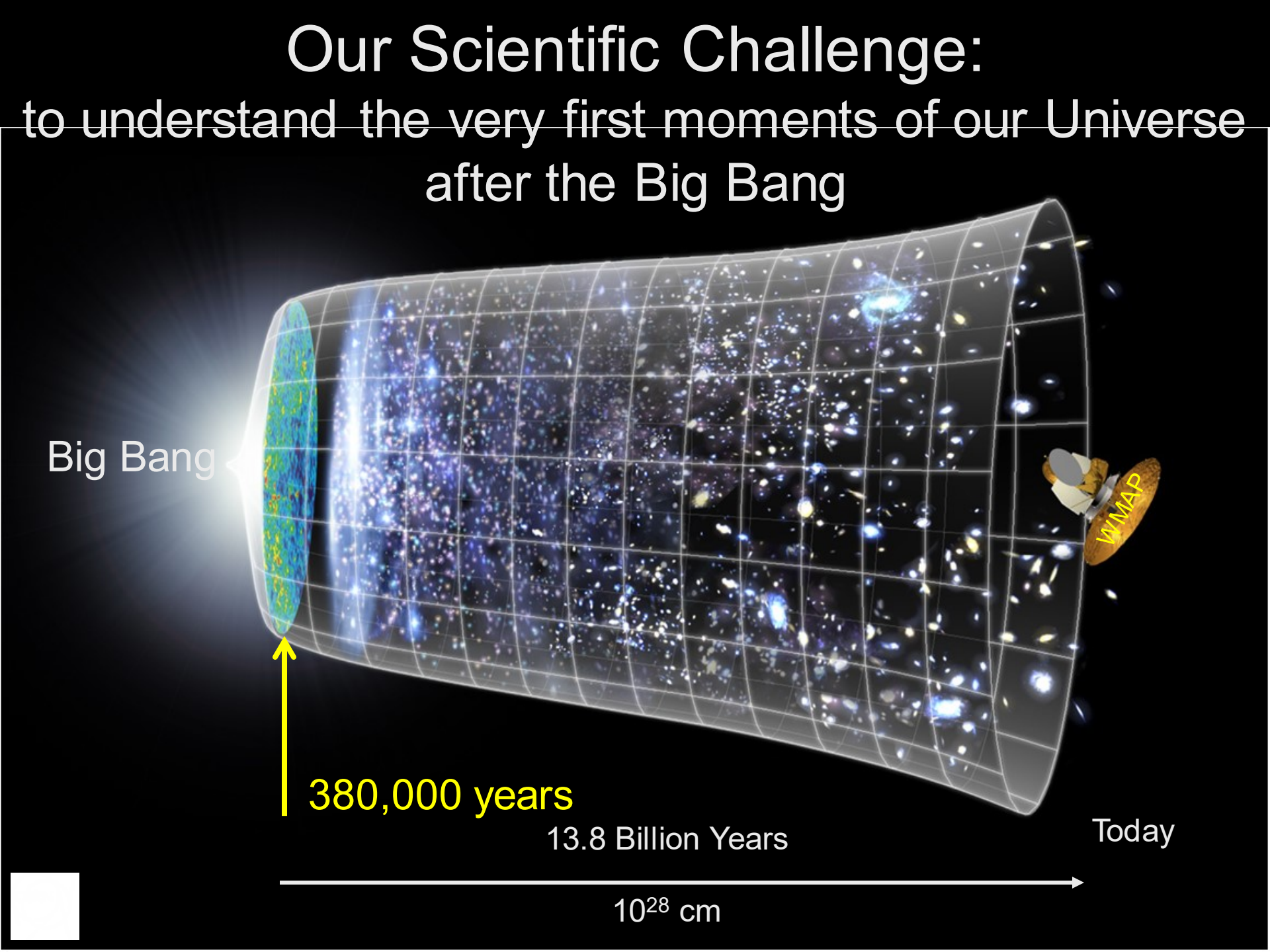


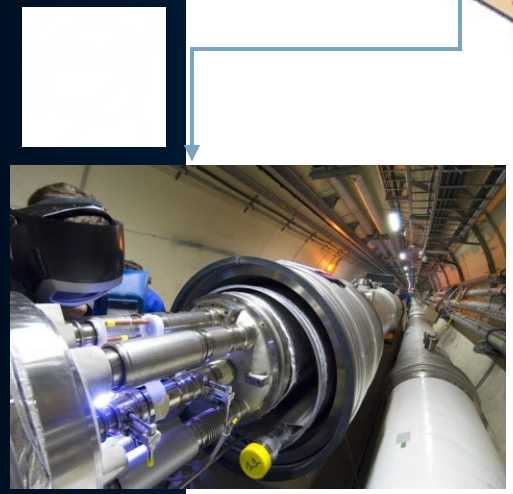


They do not all stay: where do they go?

Our Scientific Challenge:

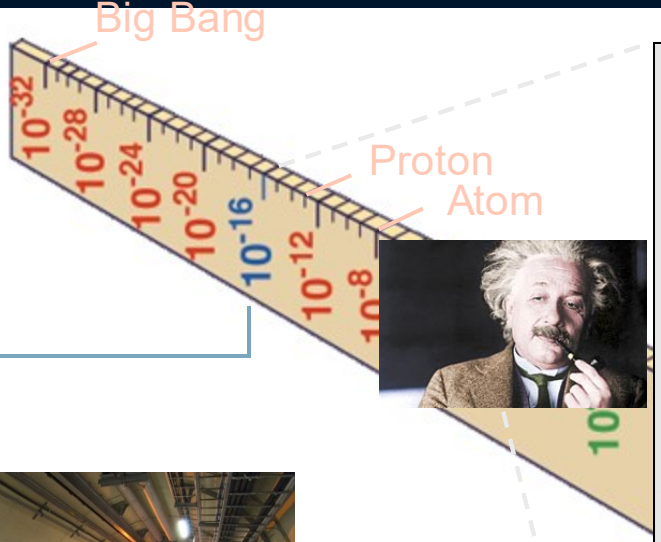
to understand the very first moments of our Universe after the Big Bang

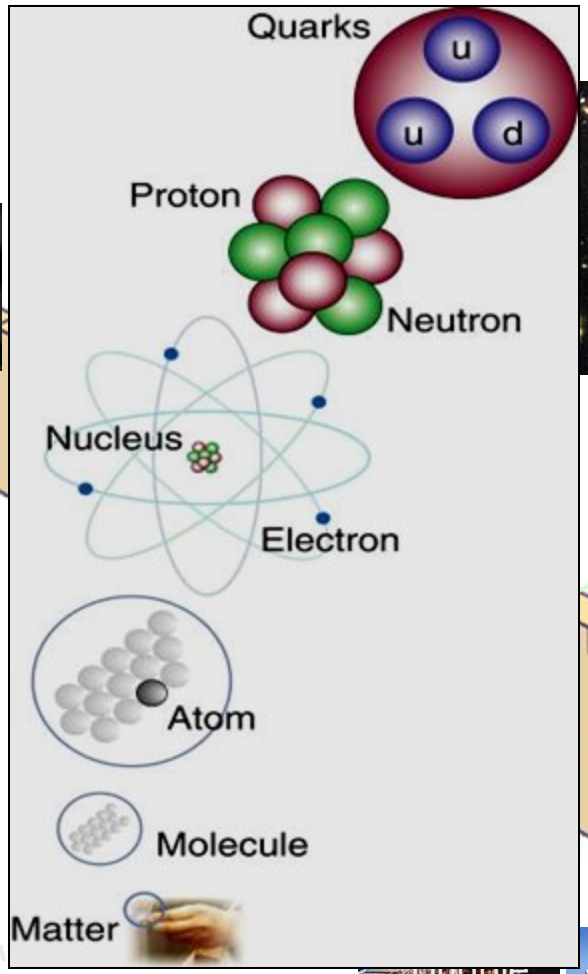




LHC

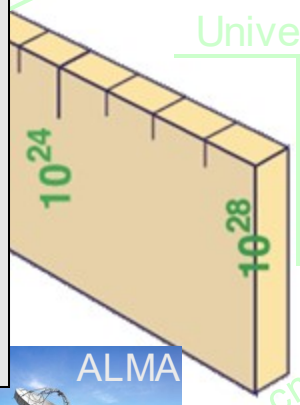
Super-Microscope





Radius of Galaxies

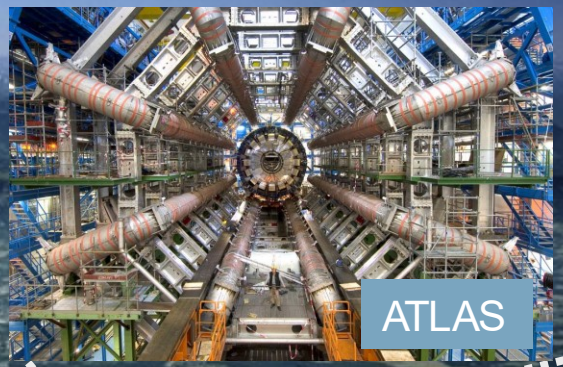
Universe

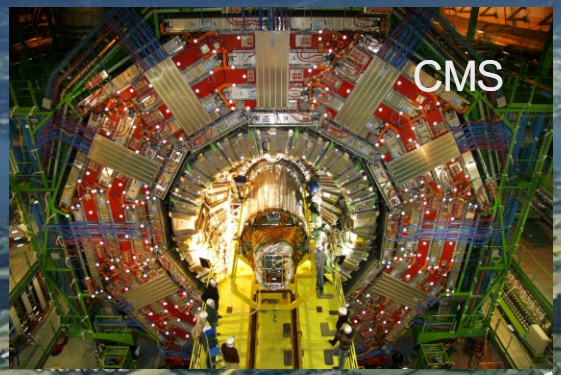


Study physics laws of first moments after Big Bang

increasing Symbiosis between Particle Physics, Astrophysics and Cosmology

2010: a New Era in Fundamental Science





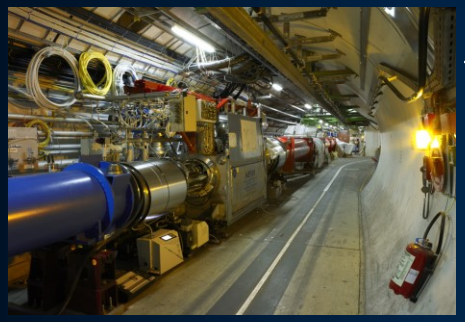
Exploration of a new energy frontier
in p-p and Pb-Pb collisions

Discovery 2012, Nobel Prize in Physics 2013

The Nobel Prize in Physics 2013 was awarded jointly to François Englert and Peter W. Higgs *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider".*

Medical Application as an Example of Particle Physics Spin-off

Combining Physics, ICT, Biology and Medicine to fight cancer

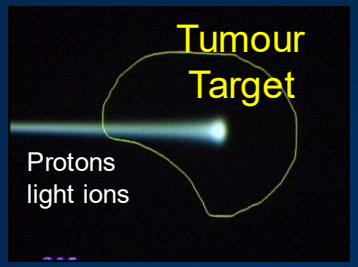


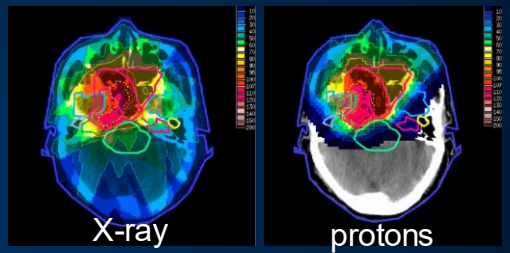
Accelerating particle beams

~30'000 accelerators worldwide

~17'000 used for medicine

Hadron Therapy





Leadership in Ion Beam Therapy now in Europe and Japan

>100'000 patients treated worldwide (45 facilities)

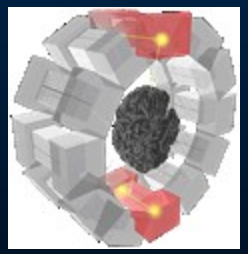
>50'000 patients treated in Europe (14 facilities)

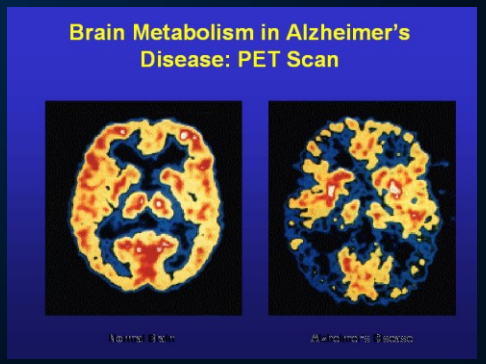
Detecting particles

Imaging

Clinical trial in Portugal, France and Italy for new breast imaging system (ClearPEM)

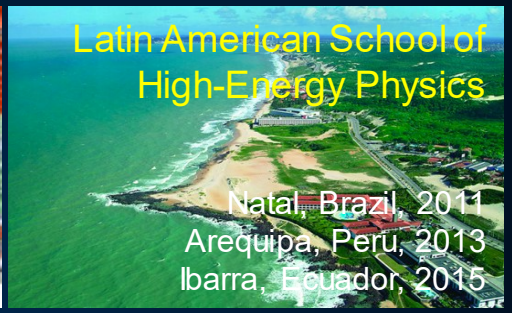
PET Scanner





CERN Education Activities

Scientists at CERN
Academic Training Programme



Young Researchers
CERN School of High Energy Physics
CERN School of Computing
CERN Accelerator School

Physics Students
Summer Students Programme

CERN Teacher Schools
International and National Programmes

CERN openlab Summer Student Programme

The CERN openlab Educational Programme

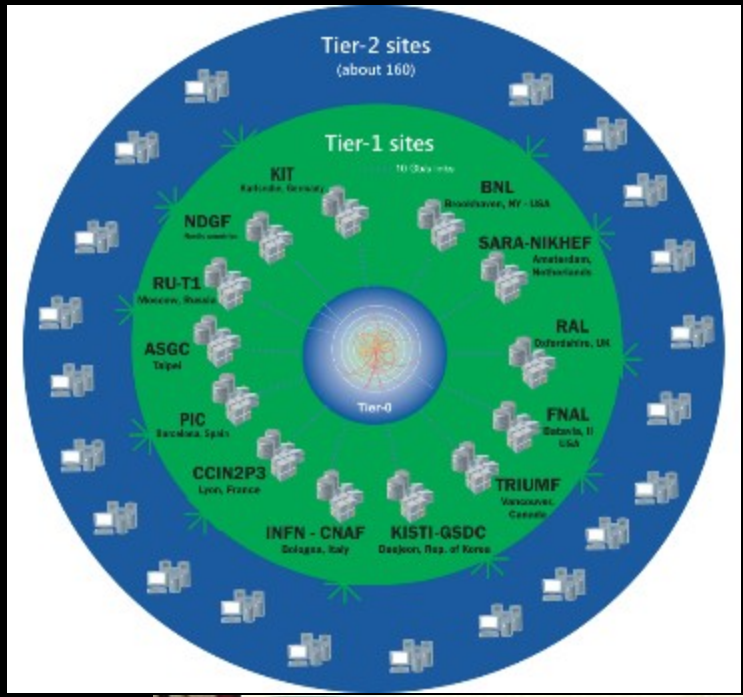
Most of the dedicated personnel in CERN openlab are young, talented Fellows receiving hands-on experience on new technologies

A comprehensive offer of general and specific workshops, training events and initiatives

Experts from industry and laboratories give lectures at events inside and outside CERN

WLCG, CERN IT & Big Data

The Worldwide LHC Computing Grid



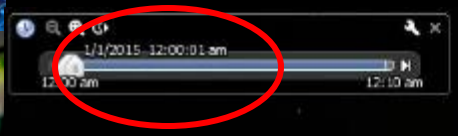
- Our task was to make use of the resources available to us – no matter where they are located

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

Tier-1 (12 centres):
- Permanent storage
- Re-processing
- Analysis

No stop for the Computing !



Running jobs: 212482
Transfer rate: 3.51 GiB/sec

Google earth

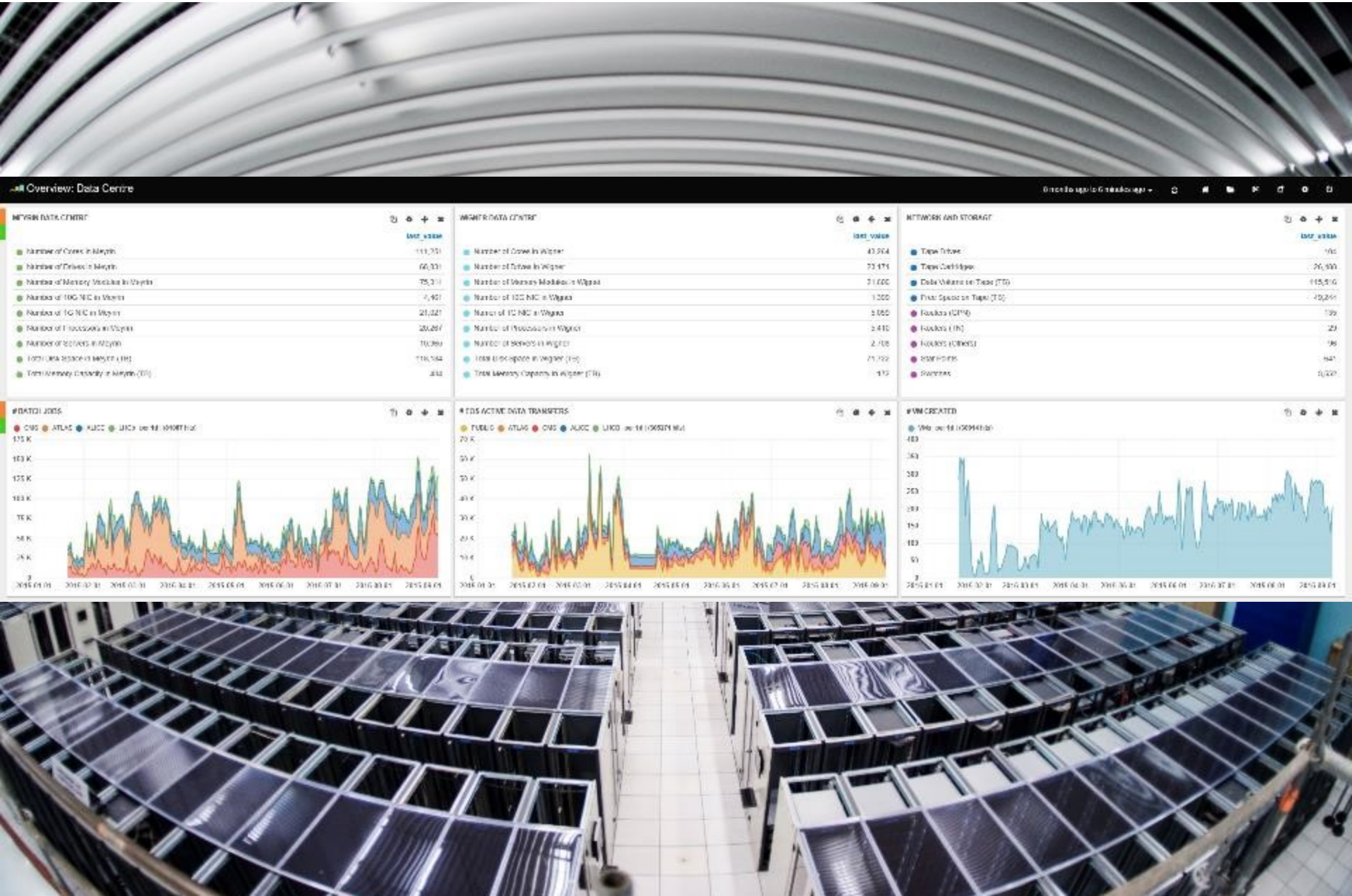
Tour Guide

29°47'59.18" N 48°48'41.91" W eye alt 19764.55 km

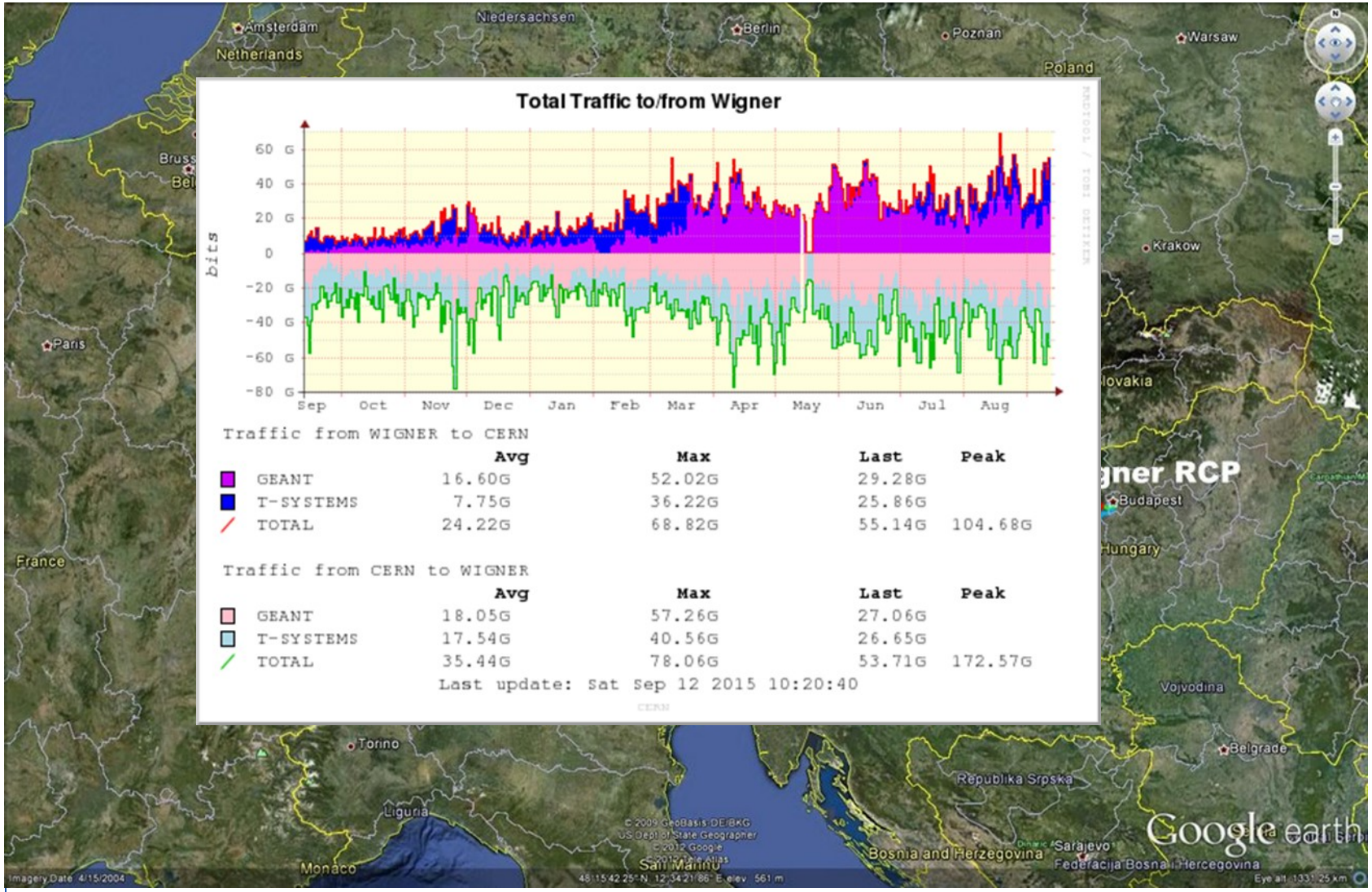
WLCG
Worldwide LHC Computing Grid

LCG

The CERN Data Centre in GVA



Linking the Data Centers



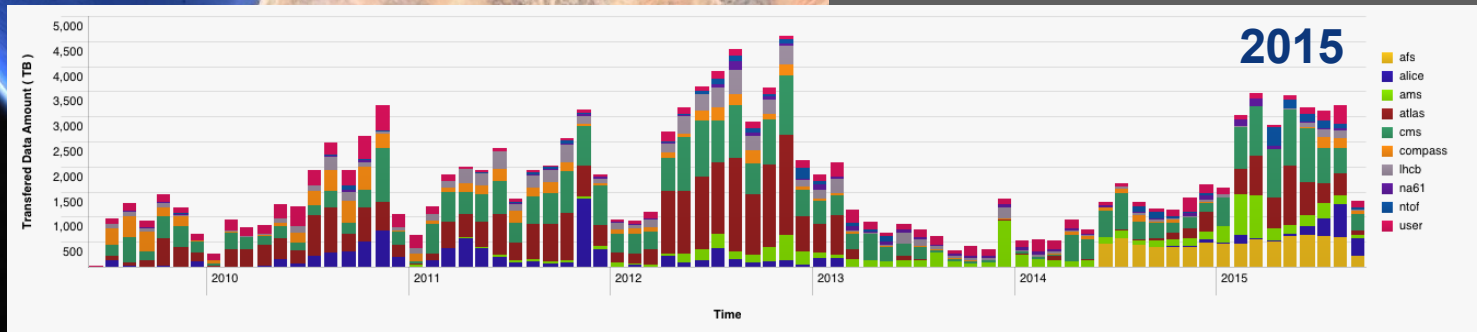
9/11/2015 10:08:41 am

Running jobs: 319053
Active CPU cores: 394939
Transfer rate: 7.61 GB/sec

WLCG fully operational with Run 2 data

Multi-core jobs

Data written to tape at Tier 0



WLCG dashboard

Tour Guide

Data Preservation

Long Term Data Preservation

- Ensure that LHC Data is preserved for future generation
  - Not only bits & bytes, also the programs that generated them
- More & more a requirement of funding agencies
- Many other disciplines, ranging from science to arts & humanities, already (very) active

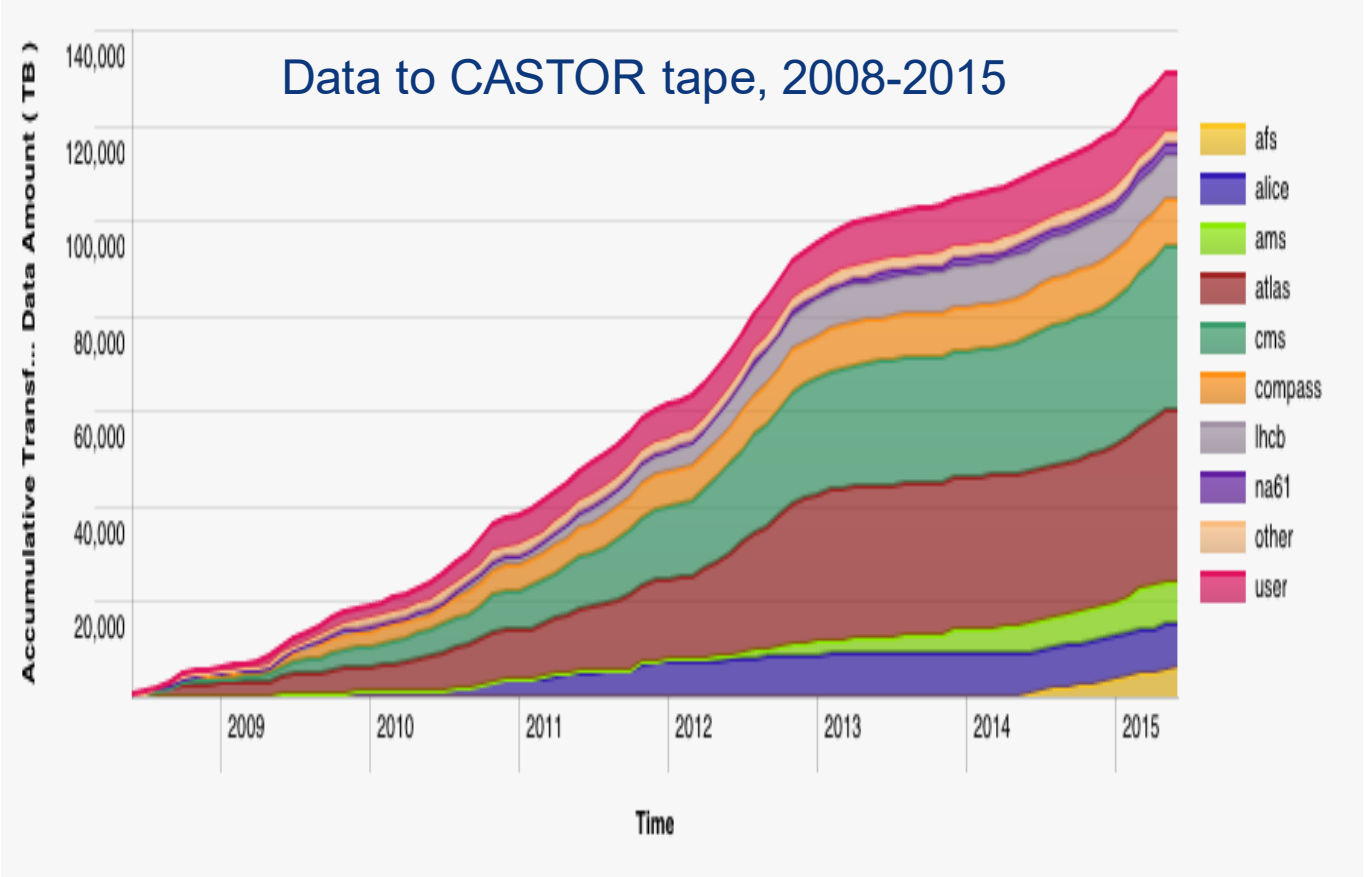
CERN Archive current numbers

Data:
- ~105 PB physics data (CASTOR)
- ~7 PB backup (TSM)

Tape libraries:
- IBM TS3500 (3+2)
- Oracle SL8500 (4)

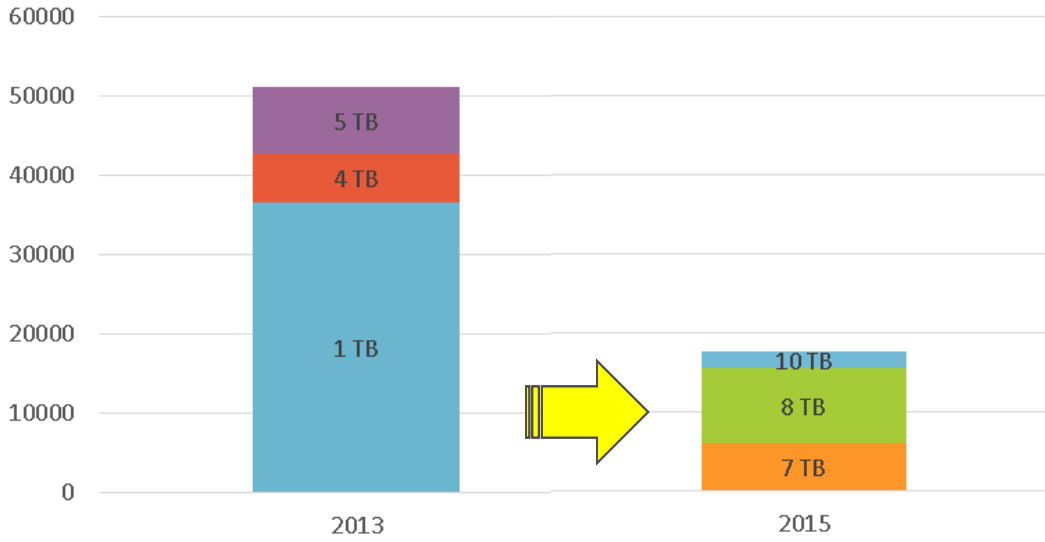
Tape drives:
- ~100 archive

Capacity:
- ~70 000 slots
- ~25 000 tapes

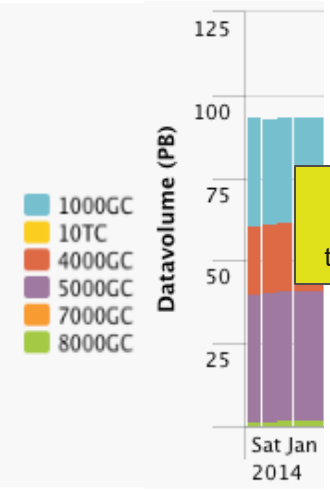


Large scale media migration

- Challenge:
  - ~85 PB of data
  - 2013: ~51 000 tapes
  - 2015: ~17 000 tapes
  - Verify all data after write
    - 3x (255PB!) pumped through the infrastructure (read->write->read)
  - Liberate library slots for new cartridges
    - Decommission ~35 000 obsolete tape cartridges
- Constraints:
  - Be transparent for user/experiment activities
  - Preserve temporal collocation
  - Finish before LHC run 2 start



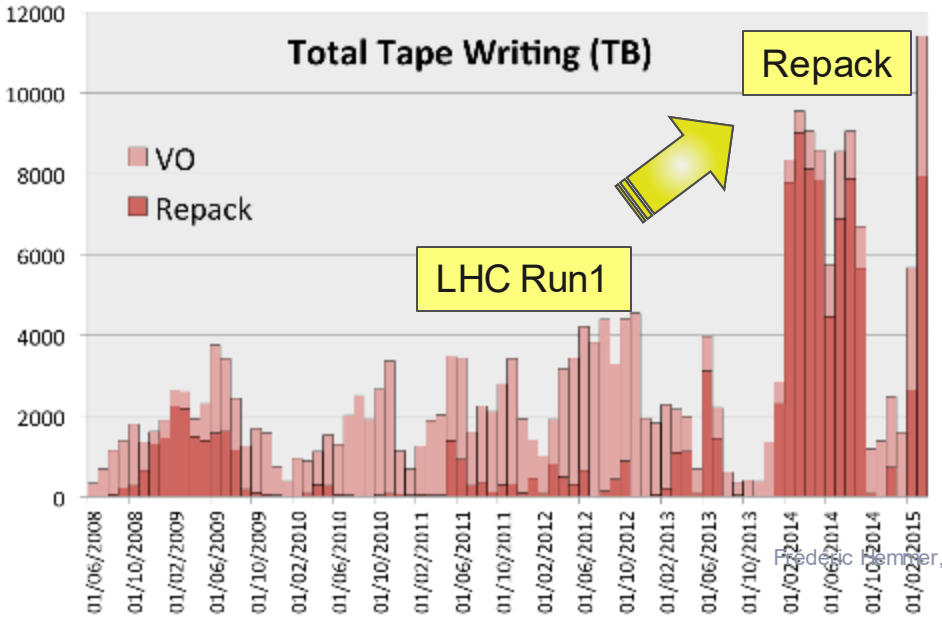
Large media migration: Repack

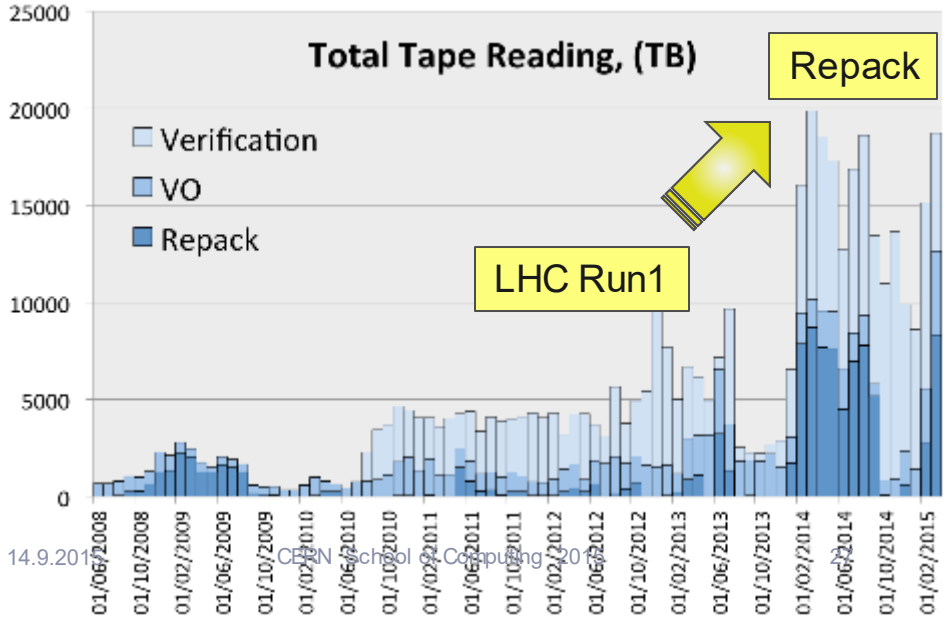


Part 1:
Oracle 5->8TB
then empty 1TB

Part 2:
IBM 4->7TB
then 1TB

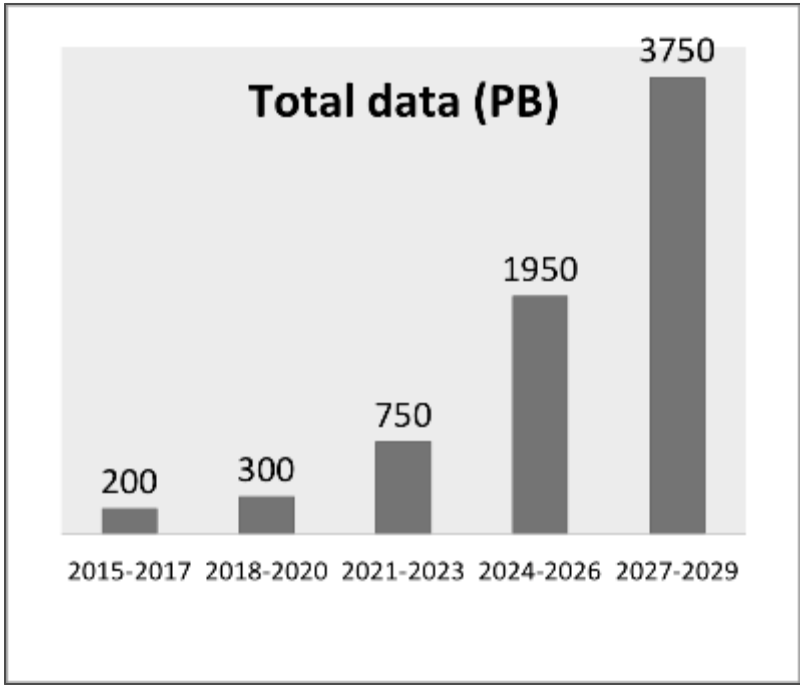
Completed !

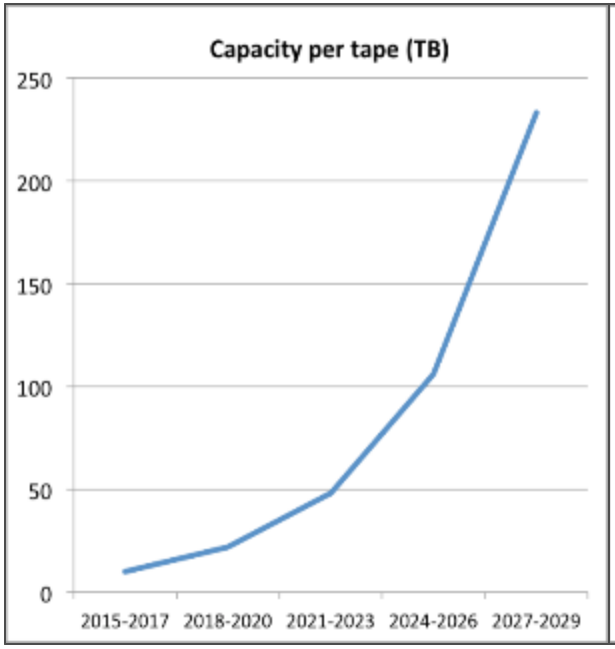


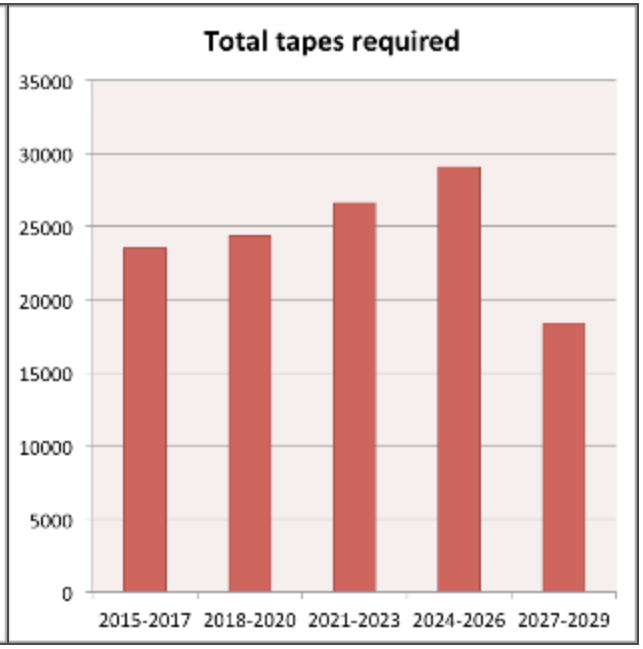


Future…

- Run-2 (2015-2018): Expecting ~50PB/year of new data (LHC + non-LHC)
  - +7K tapes / year (~35'000 free library slots)
- Run-3 (-2022): ~150PB/year. Run-4 (2023 onwards): 600PB/year..
- .. tape technology grows faster
  - tape roadmaps at 30% CAGR for at least 10 years
  - demo for 220TB tape by IBM/Fujifilm in April
- … but: market evolution is difficult to predict
  - Tape media: monopoly in shrinking market
  - disk: "duopoly"
  - Cloud storage solutions
  - Disk capacity slowdown (HAMR) .. may slowdown tape products!
  - storage slowdown == higher archiving costs

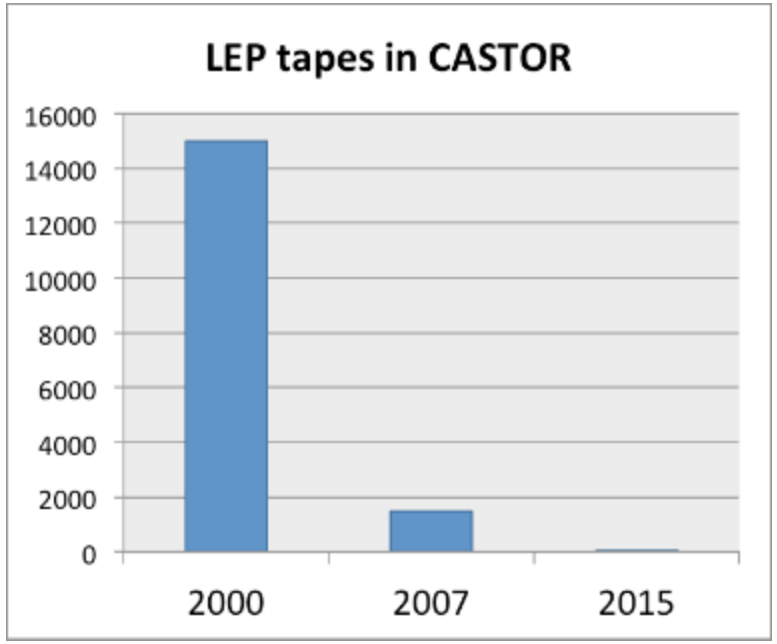






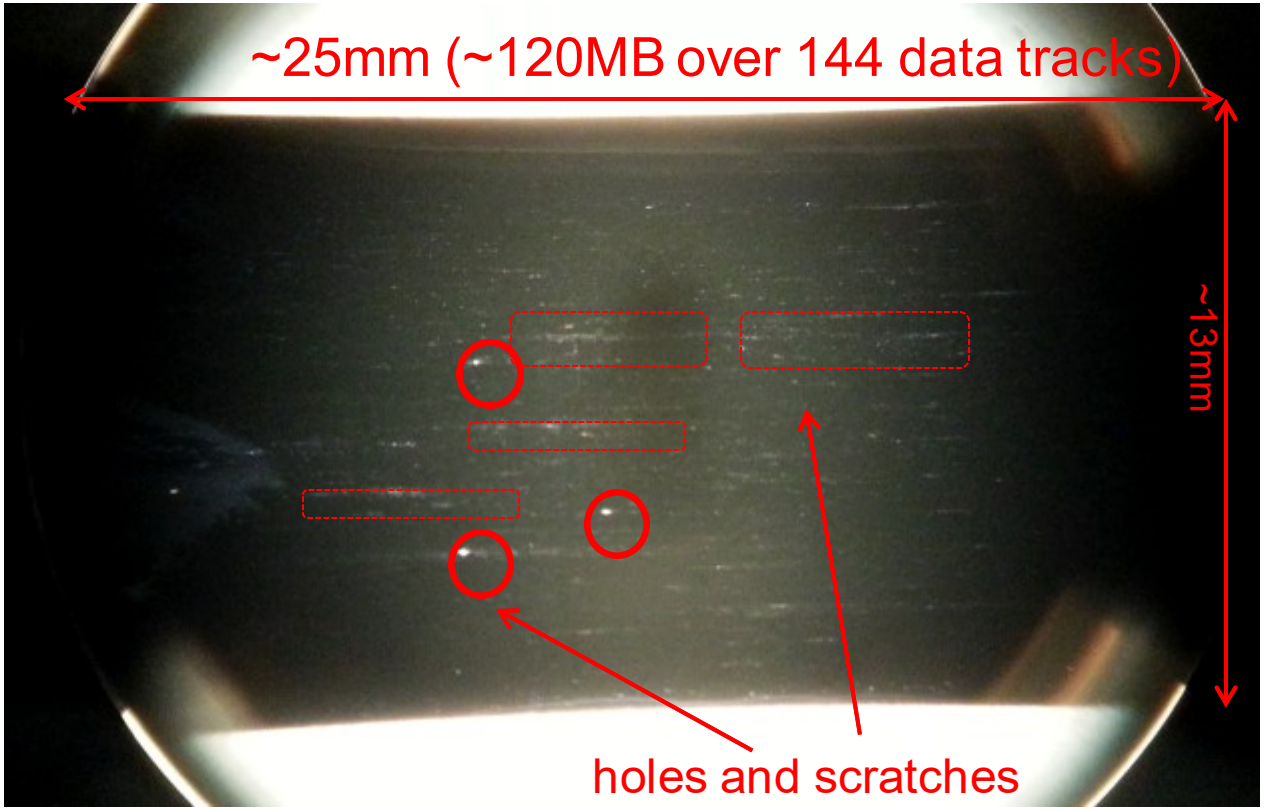
… and the past

- LEP-era data: ~370TB
- 2000:
  - ~ 15’000 tapes
- 2007:
  - ~ 1500 tapes
- 2015:
  - 30 tapes… x 2 (replicated in separate buildings)
  - Cost:



Tape contamination incident

- Identified 13 tapes in one library affected by concrete (or foam) particles
- Isolated incident by verifying all other tapes in the building
- Recovered 94% files with custom low-level tools and vendor recovery; 113 files lost



Airflows in tape libraries

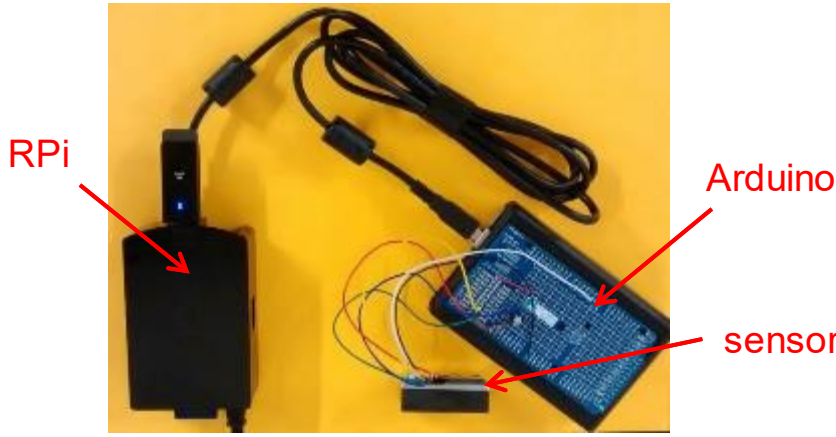
- (Our) tape libraries are not sealed nor filtered
- Over 30m^3/min of airflows per library
  - (Home vacuum cleaner: ~2m^3/min)
  - On top of already existing strong CC airflows
- Operating environment required for new-generation drives: ISO-14644 Class 8 (particles / m^3):

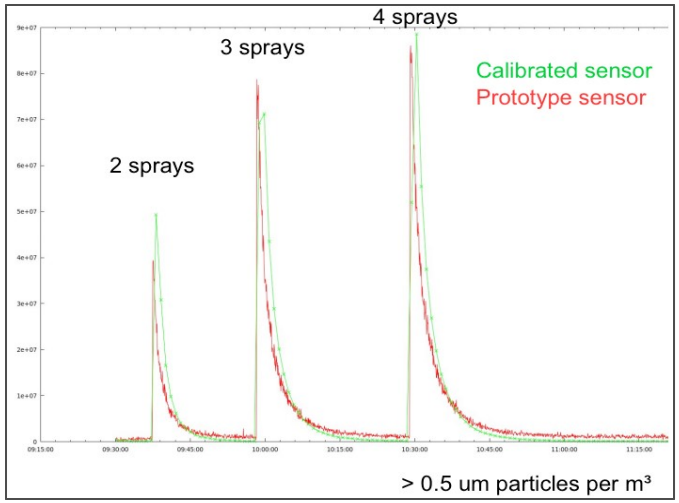
Class	>0.5 um	>1 um	>5 um
8	3 520 000	832 000	29 300

- Environmental sensitivity will continue increasing with newer drives as tape bit density grows exponentially

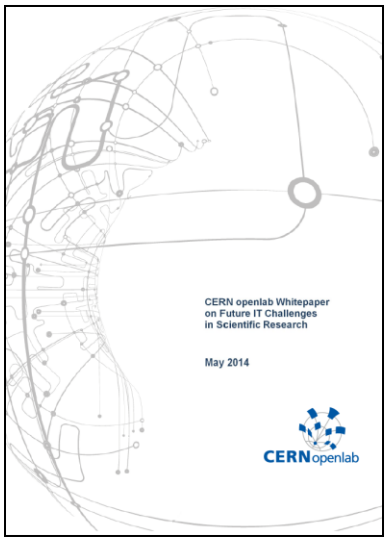
Environmental protection

- Fruitful exchanges with other HEP tape sites on CC protective measures (access and activity restrictions, special clothing, air filters etc)
- Sampling by external company and corrective actions taken at CERN-CC (air filters)
- Library cleaning by specialist company in June
- Prototyped a set of environmental sensors to be installed inside libraries, using cheap commodity components, achieving industrial precision and reaction time
  - Measure+correlate dust, temperature, humidity
  - Raise alert in case of anomalies
  - Can be integrated inside libraries
  - Done in coordination with vendor, potential for built-in solutions
  - Details: HEPiX Spring 2015 presentation





Further Reading…



http://zenodo.org/record/8765

Future IT Challenges in Scientific Research

Compute Management & Provisioning

- Data Acquisition
- Computing Platforms
- Data Storage Architectures
- Compute Management and Provisioning
- Networks and Connectivity
- Data Analytics

Update of the Computing Models of the WLCG and the LHC Experiments

http://cds.cern.ch/record/1695401

E-Infrastructure for the 21st century

http://zenodo.org/record/7592

CERN School of Computing 2015

14 September to 25 September 2015 in Kavala, Greece

Organized in collaboration with the Eastern Macedonia and Thrace Institute of Technology, Kavala, Greece

Director
Alberto Pace, CERN

Local Organising Committee Chair
Lykourgos Magafas
Dimitrios Bandekas

Technical Manager
Giuseppe Lo Presti, CERN

School Administrator
Yasemin Hauser, CERN

E-mail: computing.school@cern.ch

Deadline for Application
1st May 2015

Base Technologies
Data Technologies
Physics Computing

Lecturers

François Fluckiger, CERN, Geneva, Switzerland
Benedikt Hegner, CERN, Geneva, Switzerland
Robert G. Jacobsen, University of California at Berkeley, USA
Sebastian Lopienski, CERN, Geneva, Switzerland
Andrzej Nowak, CERN, Geneva, Switzerland
Alberto Pace, CERN, Geneva, Switzerland
Andreas J. Peters, CERN, Geneva, Switzerland
Danilo Piparo, CERN, Geneva, Switzerland
Ivica Puljak, University of Split, Croatia
Arnulf Quadt, Universität Göttingen, Germany
Benjamin Radburn-Smith, Purdue University, West Lafayette, USA

www.cern.ch/CSC


CERN