
CERN

Big Data Analytics:

Unlocking the full potential of the CERN's Large Hadron Collider.

Manuel Martín Márquez

CERN

- CERN – **European Laboratory for Particle Physics**
- Founded in **1954** by **12 Countries** for fundamental physics research in a post-war Europe
  - Major milestone in the post-World War II recovery/reconstruction process

CERN openlab

- Public-private partnership between CERN and leading ICT companies
- Accelerate cutting-edge solutions to be used by the worldwide LHC community
- Train the next generation of top engineers and scientists.

Partners

HUAWEI

intel

ORACLE

SIEMENS

Contributors

rackspace
the open cloud company

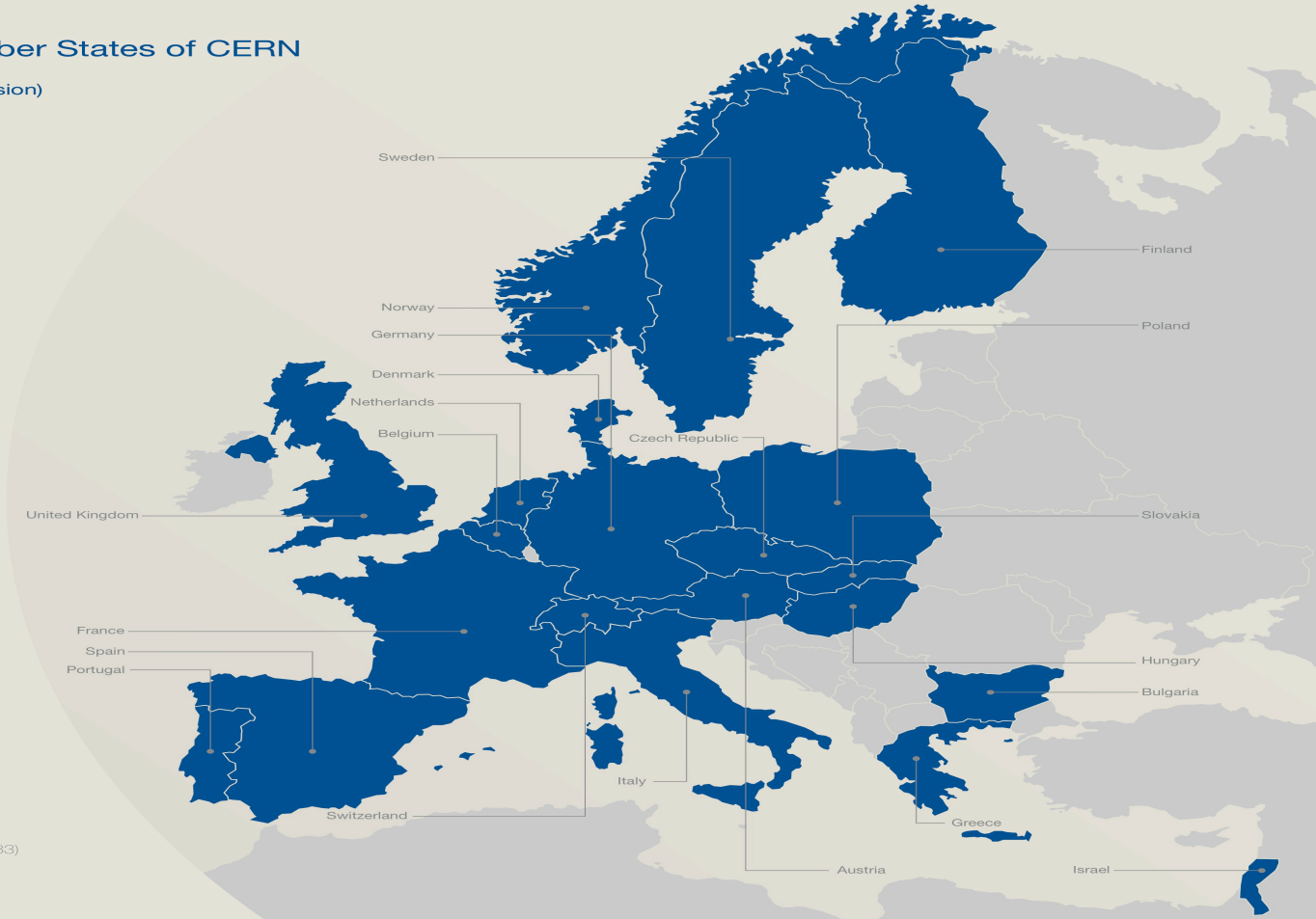
Associates

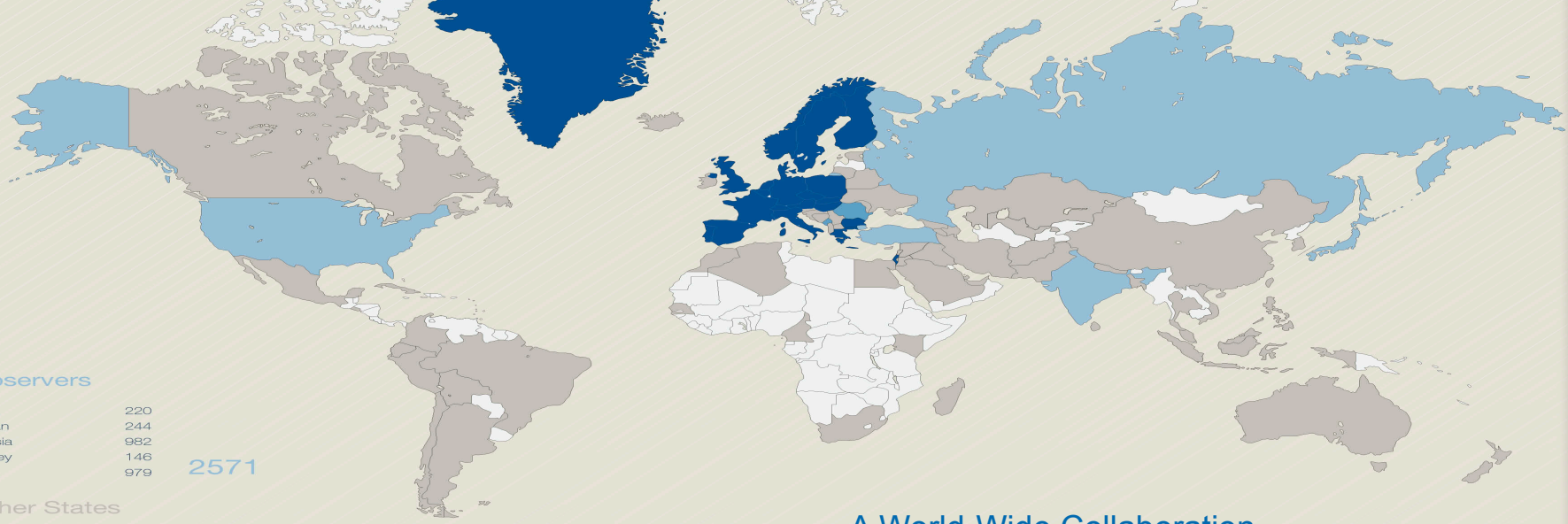
Yandex

The twenty one Member States of CERN

Member States (Dates of accession)

- Austria (1959)
- Belgium (1953)
- Bulgaria (1999)
- Czech Republic (1993)
- Denmark (1953)
- Finland (1991)
- France (1953)
- Germany (1953)
- Greece (1953)
- Hungary (1992)
- Israel (2014)
- Italy (1953)
- Netherlands (1953)
- Norway (1953)
- Poland (1991)
- Portugal (1986)
- Slovakia (1993)
- Spain (1/1961-12/1968-1/1983)
- Sweden (1953)
- Switzerland (1953)
- United Kingdom (1953)





Observers

India	220
Japan	244
Russia	982
Turkey	146
USA	979

2571

Other States

Afghanistan	1	El Salvador	1	Pakistan	41
Albania	2	Estonia	16	Palestine (O.T.).	4
Algeria	8	Georgia	36	Peru	8
Argentina	11	Gibraltar	1	Philippines	1
Armenia	25	Hong Kong	1	Saudi Arabia	3
Australia	25	Iceland	4	Senegal	1
Azerbaijan	8	Indonesia	1	Singapore	2
Bangladesh	4	Iran	28	Sint Maarten	2
Belarus	47	Ireland	22	Slovenia	27
Bolivia	3	Jordan	2	South Africa	16
Bosnia & Herzegovina	1	Kenya	1	Sri Lanka	5
Brazil	108	Korea, D.P.R.	1	Syria	2
Cameroon	1	Korea Rep.	117	Thailand	12
Canada	134	Kuwait	1	T.F.Y.R.O.M.	1
Cape Verde	1	Lebanon	12	Tunisia	6
Chile	12	Lithuania	19	Ukraine	55
China	280	Luxembourg	4	Uzbekistan	4
China (Tapei)	45	Madagascar	4	Venezuela	9
Colombia	30	Malaysia	15	Viet Nam	9
Croatia	35	Mauritius	1	Zimbabwe	2
Cuba	7	Mexico	64		
Cyprus	16	Montenegro	3		
Ecuador	3	Morocco	12		
Egypt	19	Nepal	5		
		New Zealand	7		

1415

A World-Wide Collaboration

Member States

Austria	99	Greece	152	Slovakia	88
Belgium	106	Hungary	68	Spain	337
Bulgaria	75	Israel	51	Sweden	75
Czech Republic	202	Italy	1686	Switzerland	180
Denmark	53	Netherlands	153	United Kingdom	640
Finland	87	Norway	61		
France	751	Poland	229		
Germany	1150	Portugal	109		

6352

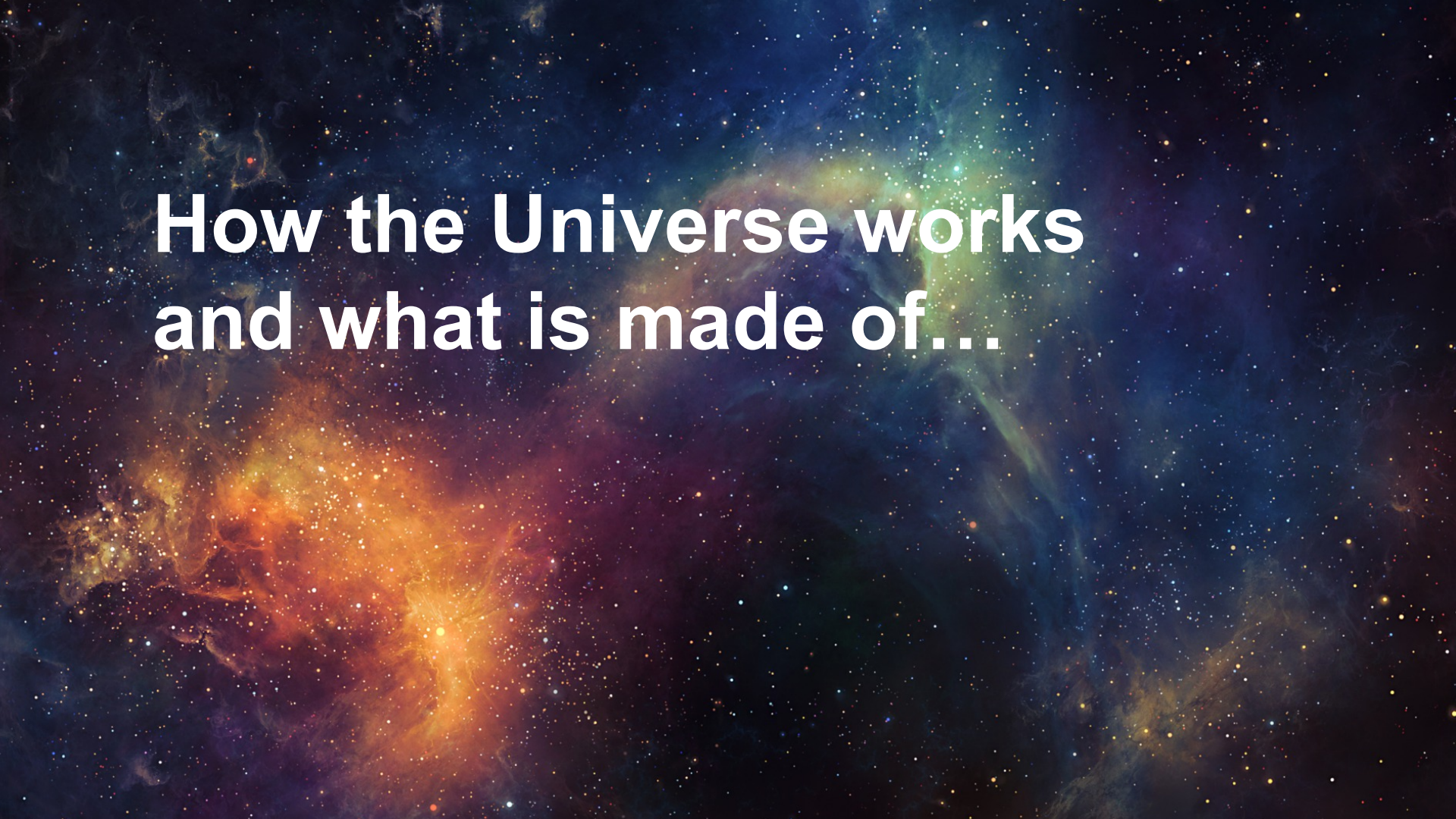
Candidate for Accession

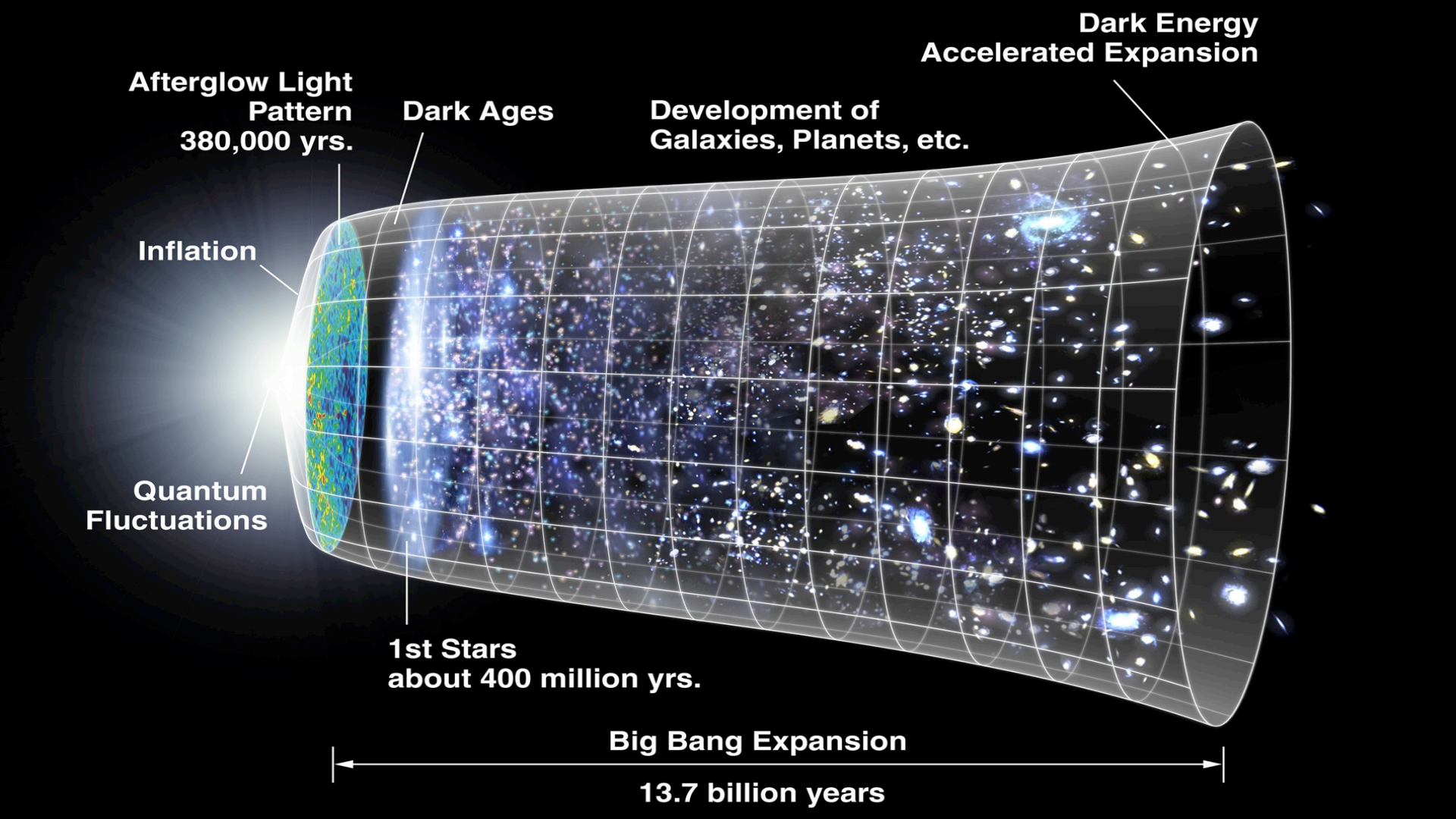
Romania	118

Associate Members in the Pre-stage to Membership

Serbia	41

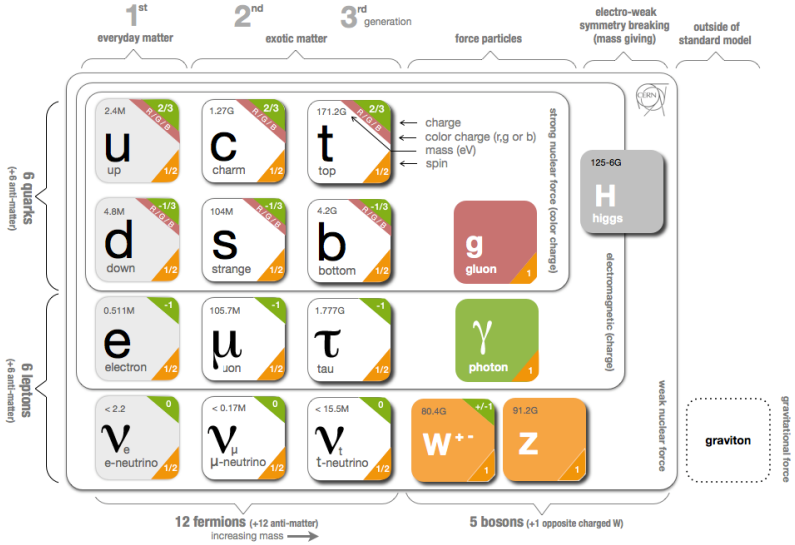
Distribution of All CERN Users by Nationality on 14 January 2014


How the Universe works
and what is made of…


Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
380,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.7 billion years

Fundamental Research

- Why do particles have mass?



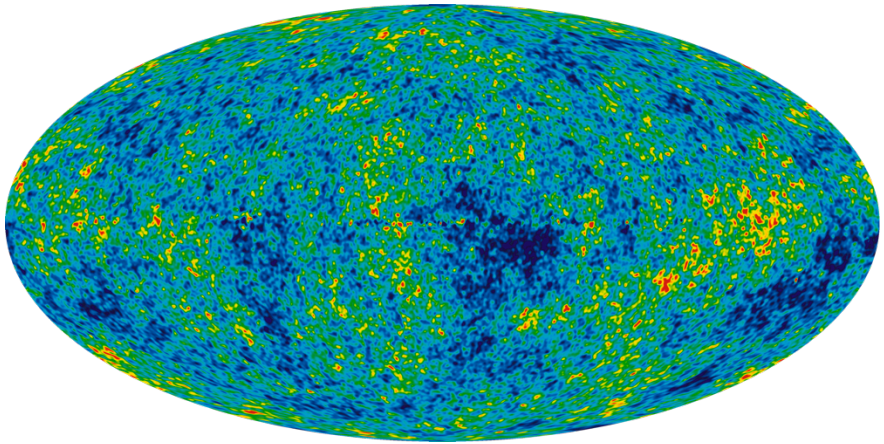
Fundamental Research

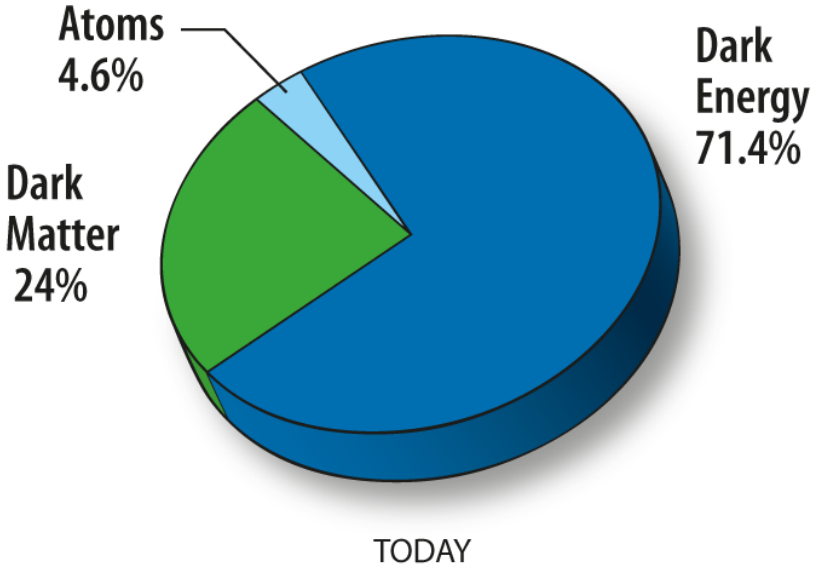
- Why is there no antimatter left in the Universe?
  - Nature should be symmetrical

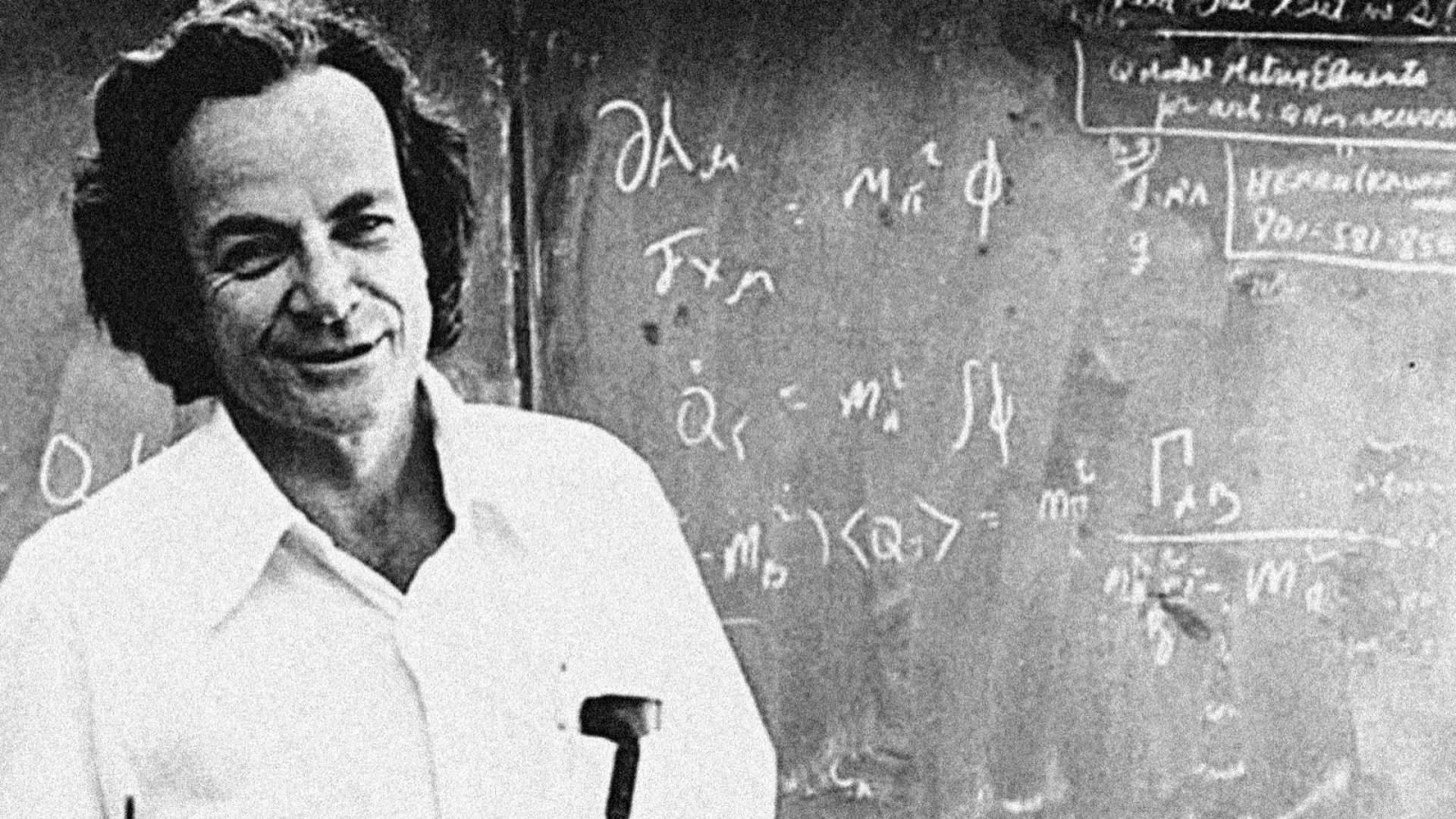
- What was matter like during the first second of the Universe, right after the "Big Bang"?
  - A journey towards the beginning of the Universe gives us deeper insight.

Fundamental Research

- What is 95% of the Universe made of?






$\partial A_\mu = M_\pi^2 \phi$
$F_{\chi\mu}$
$\partial_\mu = m_\pi^2 \int \psi$
$(m_b^2) \langle Q_5 \rangle = m_\pi^2 \frac{\Gamma_{\mu 5}}{m_\pi^2 - M^2}$

The Large Hadron Collider (LHC)

Largest machine in the world

27km, 6000+ superconducting magnets

Fastest racetrack on Earth

Protons circulate 11245 times/s (99.9999991% the speed of light)

Emptiest place in the solar system

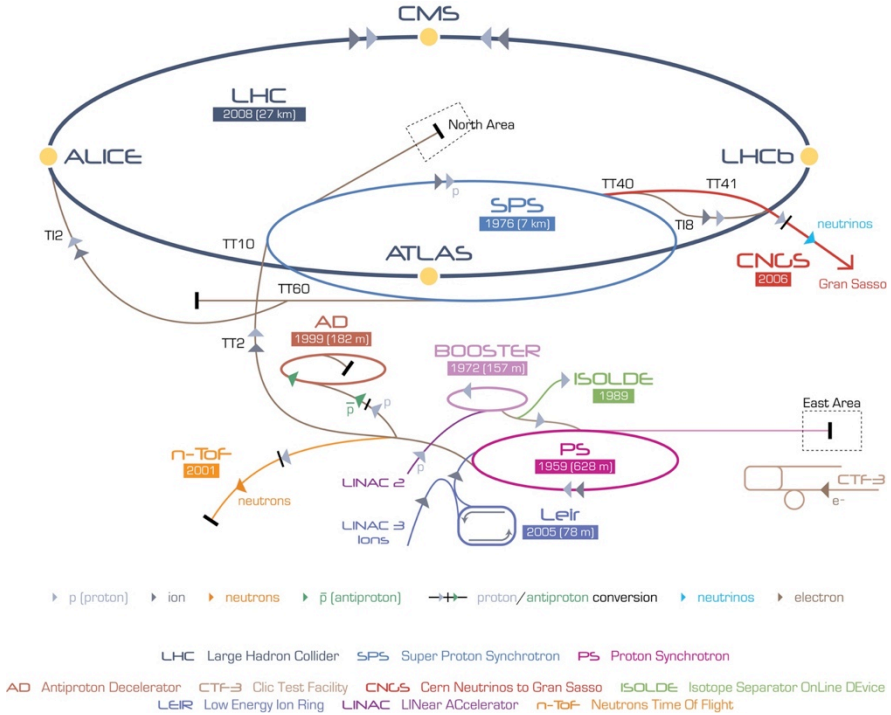
High vacuum inside the magnets

Hottest spot in the galaxy

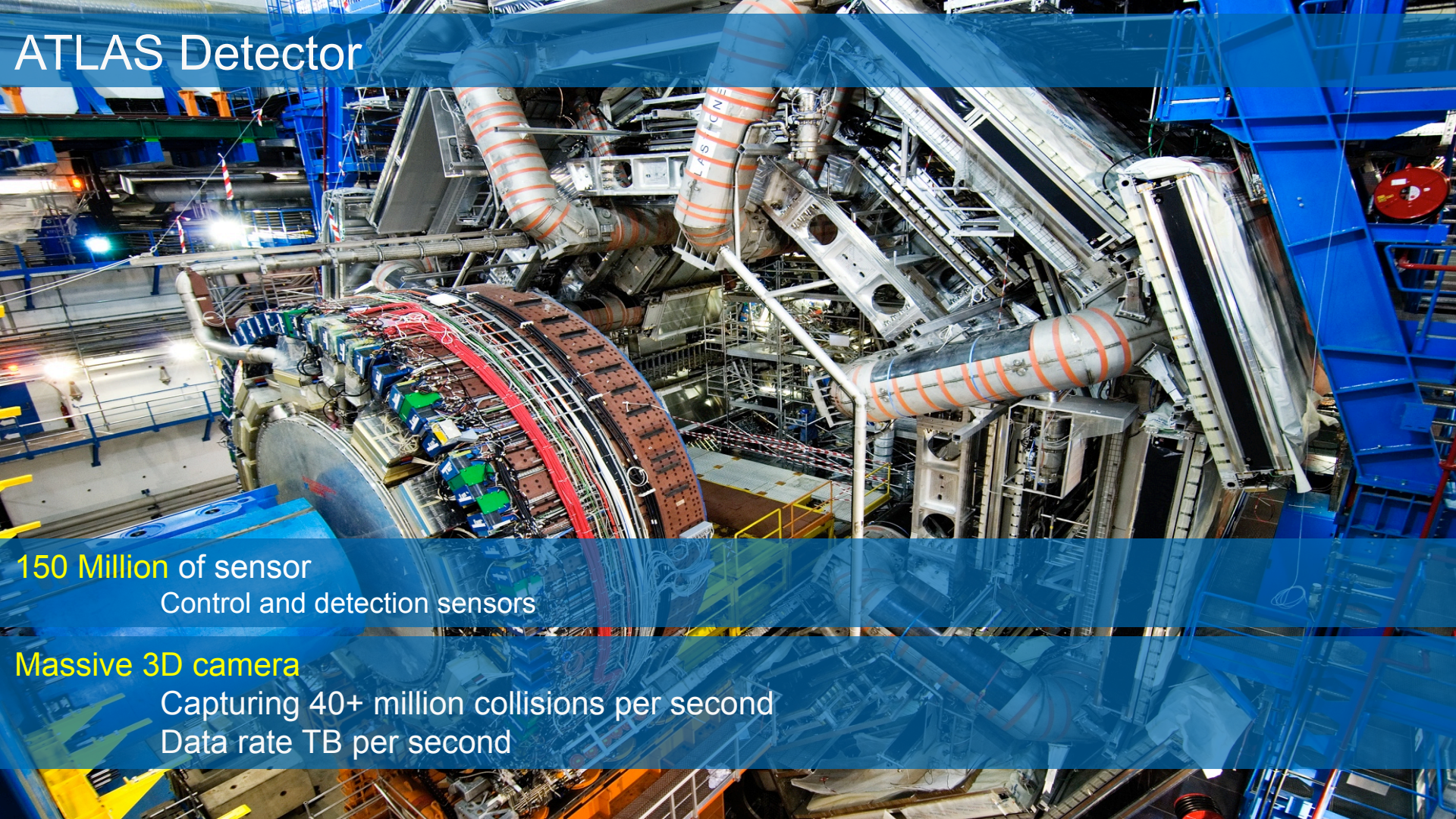
During Lead ion collisions create temperatures 100 000x hotter than the heart of the sun;

The Large Hadron Collider (LHC)

CERN's Accelerator Complex



ATLAS Detector



150 Million of sensor

Control and detection sensors

Massive 3D camera

Capturing 40+ million collisions per second

Data rate TB per second

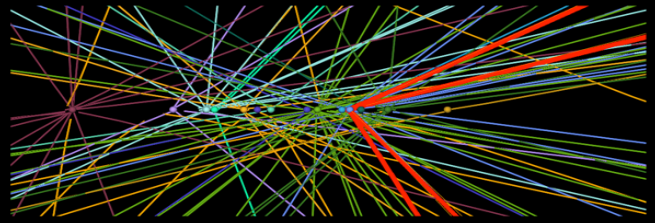
CMS Detector

Raw Data

- Was a detector element hint?
- How much energy?
- What time?

Reconstructed Data

- Particle Type
- Origin
- Momentum of tracks (4 vectors)
- Energy in cluster (jets)
- Calibration Information

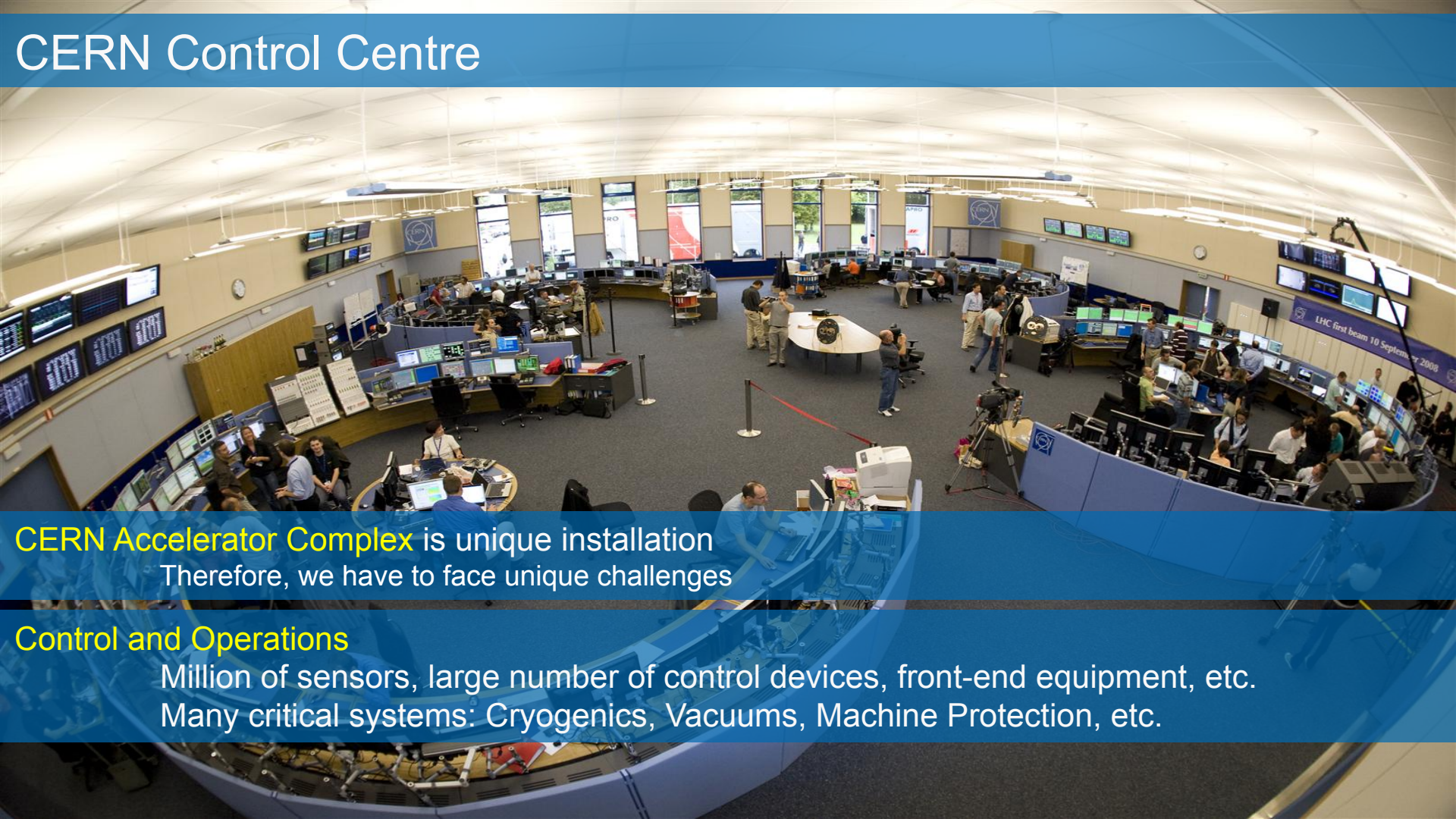


Higgs Boson Discovery

2012

Higgs to 4μ candidate event

CERN Control Centre



CERN Accelerator Complex is unique installation

Therefore, we have to face unique challenges

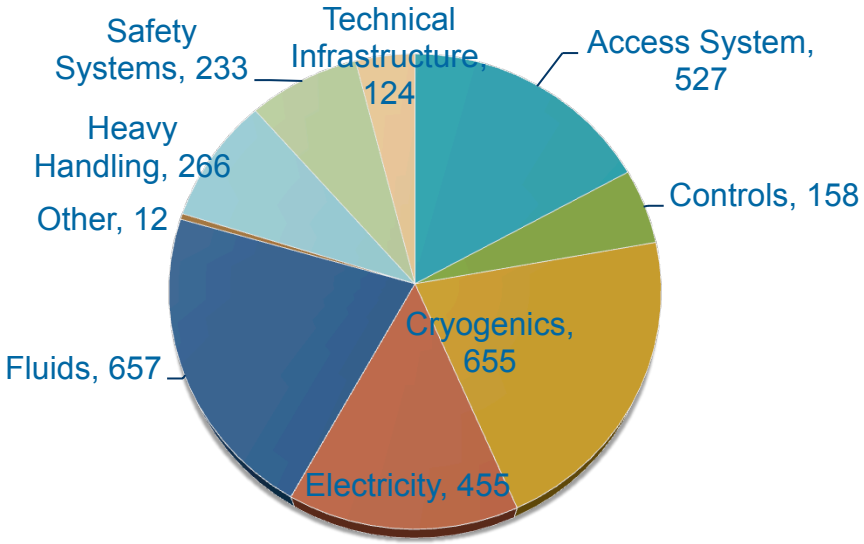
Control and Operations

Million of sensors, large number of control devices, front-end equipment, etc.

Many critical systems: Cryogenics, Vacuums, Machine Protection, etc.

Data Analytics Challenges

LHC Corrective Intervention: 3087 / year



Data Analytics Challenges

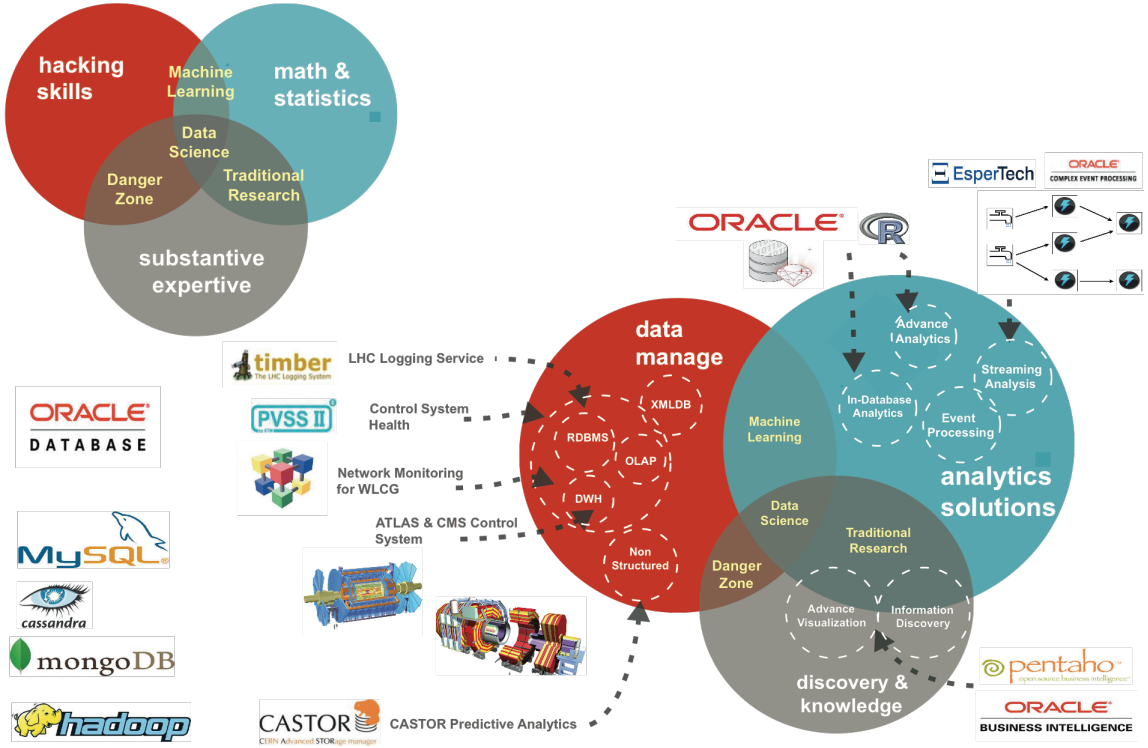
- A look into the near Future
  - LHC run 2 (2015)

				2015
Parameter	**2010**	**2011**	**2012**	**design value**
Beam energy	3.5	3.5	4	7
β^* in IP 1 and 5 (m)	2.0/3.5	1.5/1.0	0.6	0.55
Bunch spacing (ns)	150	75/50	50	25
Max. number of bunches	368	1380	1380	2808
Max. bunch intensity (protons per bunch)	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}	1.15×10^{11}
Normalized emittance at start of fill (mm mrad)	≈2.0	≈2.4	≈2.5	3.75
Peak luminosity ($cm^{-2}s^{-1}$)	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}	1×10^{34}
Max. mean number of events per bunch crossing	4	17	37	19
Stored beam energy (MJ)	≈28	≈110	≈140	362

Data Analytics Challenges

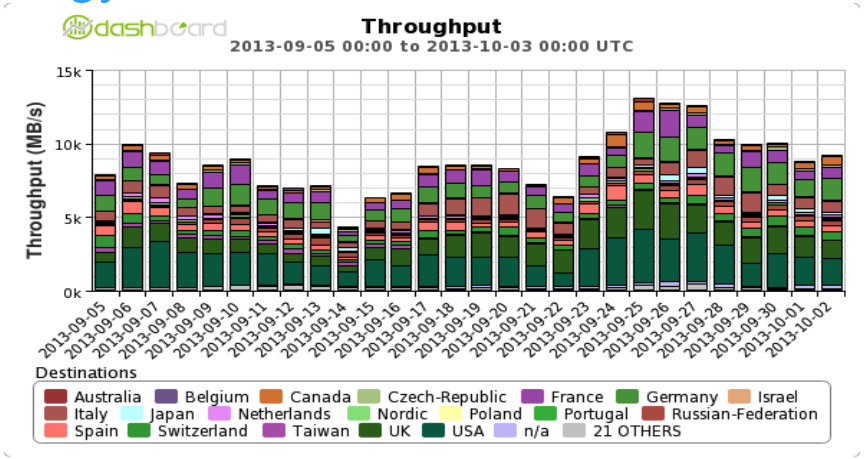
- Profit from our data investment
  - Extracting knowledge.
- Optimize our systems is mandatory
  - Reducing and predicting faults and corrective interventions
  - Increase the availability and operations efficiency
- Control and Monitoring Systems
  - Proactive
  - Predictive
  - Intelligent

Data & Analytics Environment



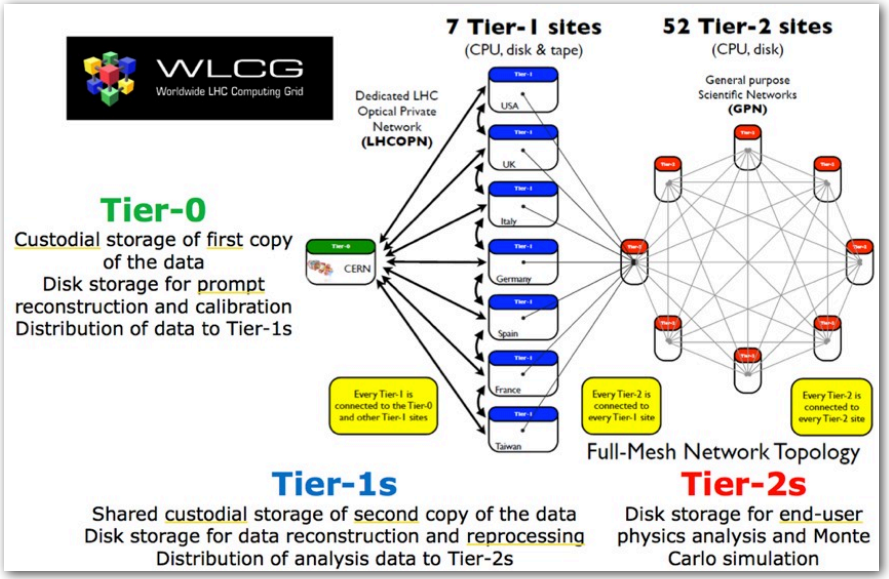
CERN's Data Analytics Use Cases

- WLCG relies heavily on the underlying networks
- Network Monitoring WLCG
  - Correlation in time and topology
  - Real-Time Analytics
    - Root Cause Analysis
    - Early warning systems



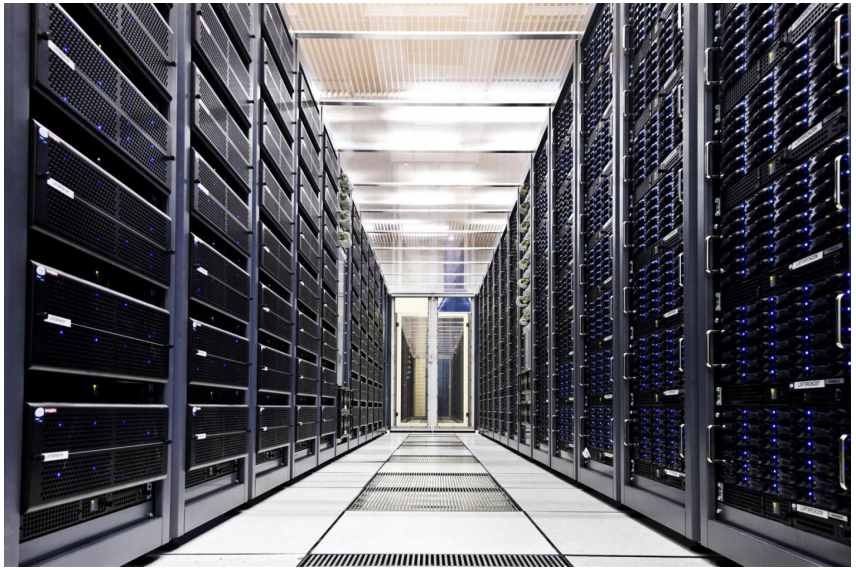
CERN's Data Analytics Use Cases

- Intelligent Data Placement for CMS
- Resources optimization
  - Minimize number of replicas
  - Remove Obsolete
  - Job time in data access
  - Job time in data analysis
- Resources Prediction



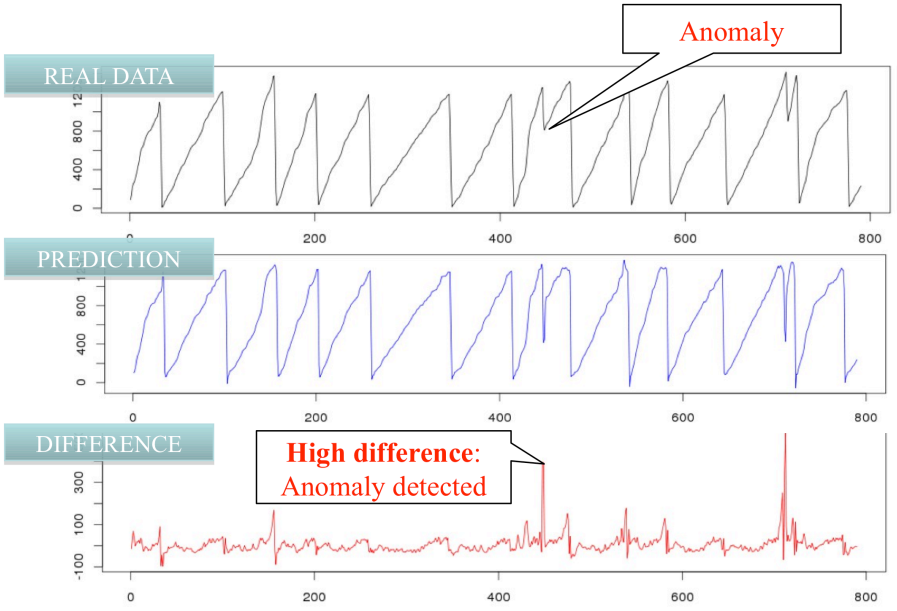
CERN's Data Analytics Use Cases

- CASTOR - CERN Advance Storage Manager
  - CERN Mass Storage Solution
    - Disk + Tapes
    - 12k disks, 30k tapes
- Expert system
  - Spot ongoing incidents
- Predictive analysis
  - Predict problem occurrences



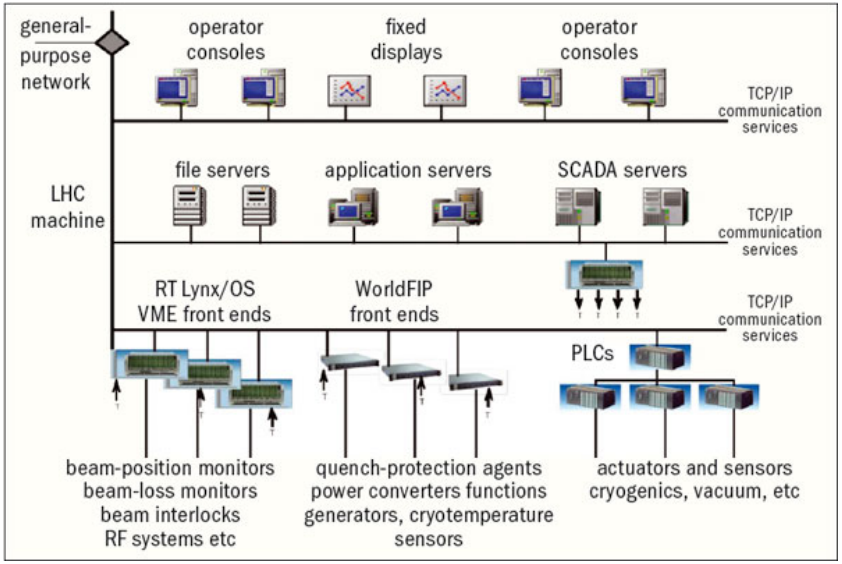
CERN's Data Analytics Use Cases

- CASTOR – Anomaly Detection
  - Data to be transferred
    - Queue data
  - Prediction Model
    - Real - Prediction



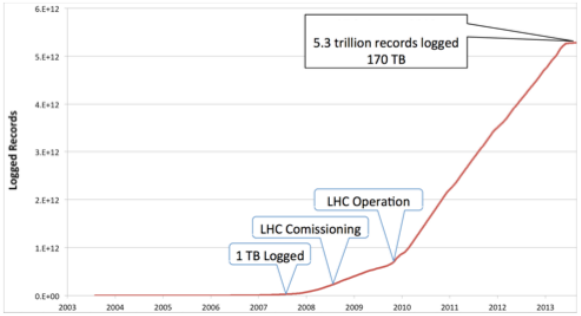
CERN's Data Analytics Use Cases

- Control Systems
  - Control system Health
    - Gas Breakdown
    - Predictive maintenance
      - Cryogenics
      - Vacuum
      - Machine Protection
  - Quench Detection

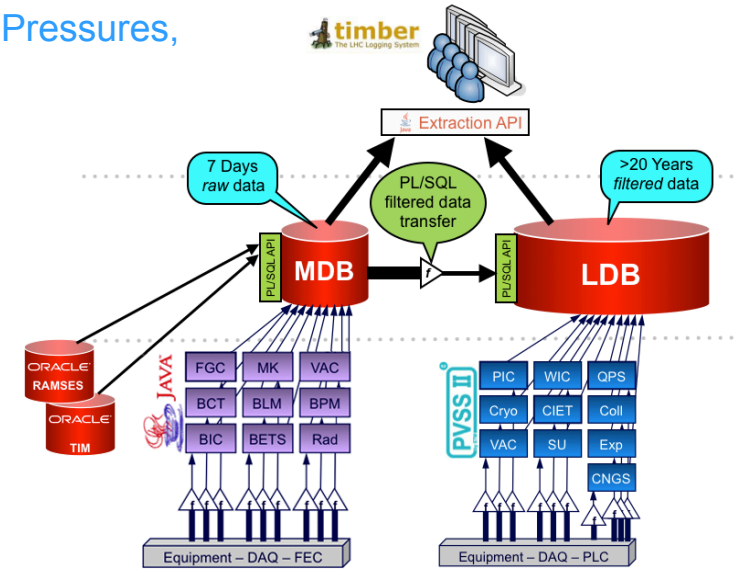


CERN Accelerators Control System

- Close to 1 million pre-defined signals
  - Cryogenics temperatures,
  - Magnetic field strengths, Power dissipation, Vacuum Pressures,
  - Beam intensities and positions…etc…
- About 5 million daily/average data requests
- Throughput over 100TB/Year, 300TB in 2015

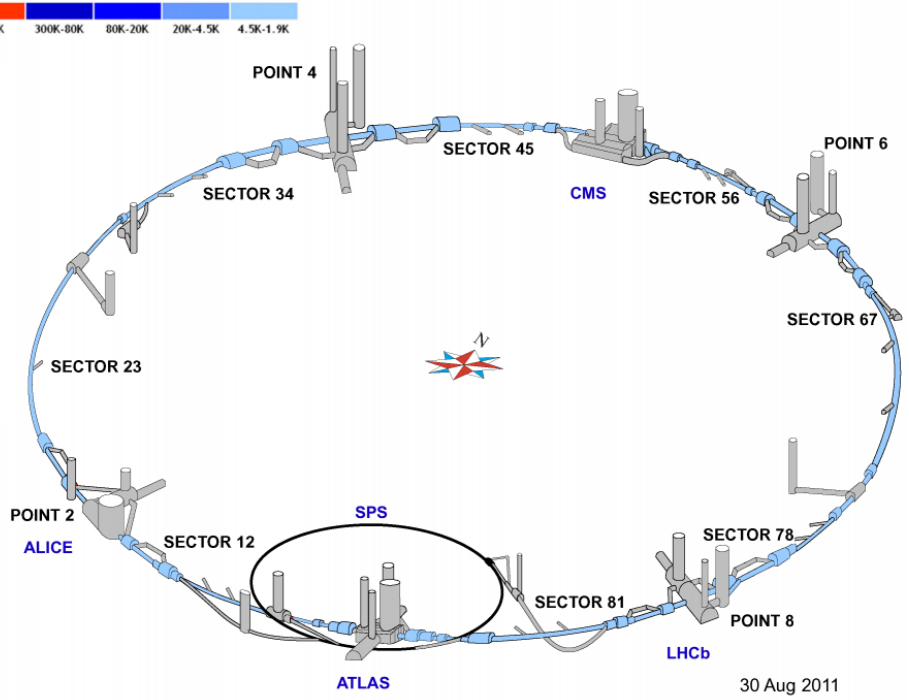


Evolution of logged records.



Largest Cryogenics Installation

Instruments	Range	Total
TT (temperature)	1.6- 300K	9500
PT (pressure)	0-20 bar	2200
LT (level)	Various	540
EH (heaters)	Various	2500
CV (Control Valves)	0 - 100 %	3800
PV/QV (On Off Valves)	--	2000



Largest Cryogenics Installation

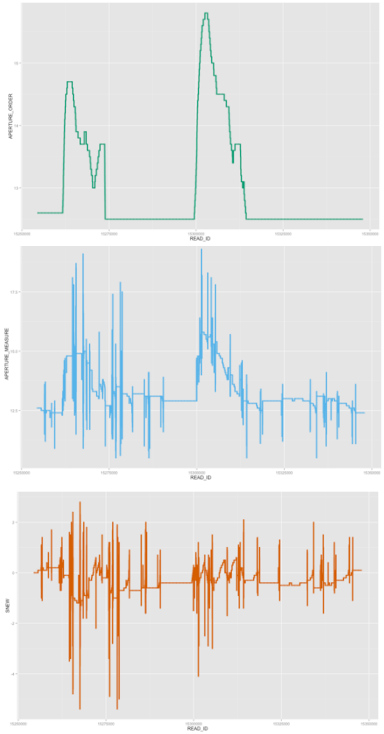
- Study based on sectors
  - L4, R4, L8 and R8.
- Sensor Outputs
  - aperture order (%)
  - aperture measured (%)
- Three different status:
  - Faulty,
  - Not faulty
  - Unknown

Largest Cryogenics Installation

- Signals used:
  - S = aperture order - aperture measured
- Features extractions based on S
  - Variance
  - Percentile 99.9
  - Rope distance – R(S)
  - Noise Band – B(S)

$$R(S) = \frac{1}{N}\sum_{i=2}^{N}\left|S(i) - S(i-1)\right|$$

$$B(S) = \frac{\left|\sum_{k=1}^{N_{fft}/2} Pxx(k)\right|^2}{\sum_{k=1}^{N_{fft}/2} Pxx(k)^2}$$



- **Automatic Faulty Valves Detection System**
  - SVM - Support Vector Machine

www.cern.ch