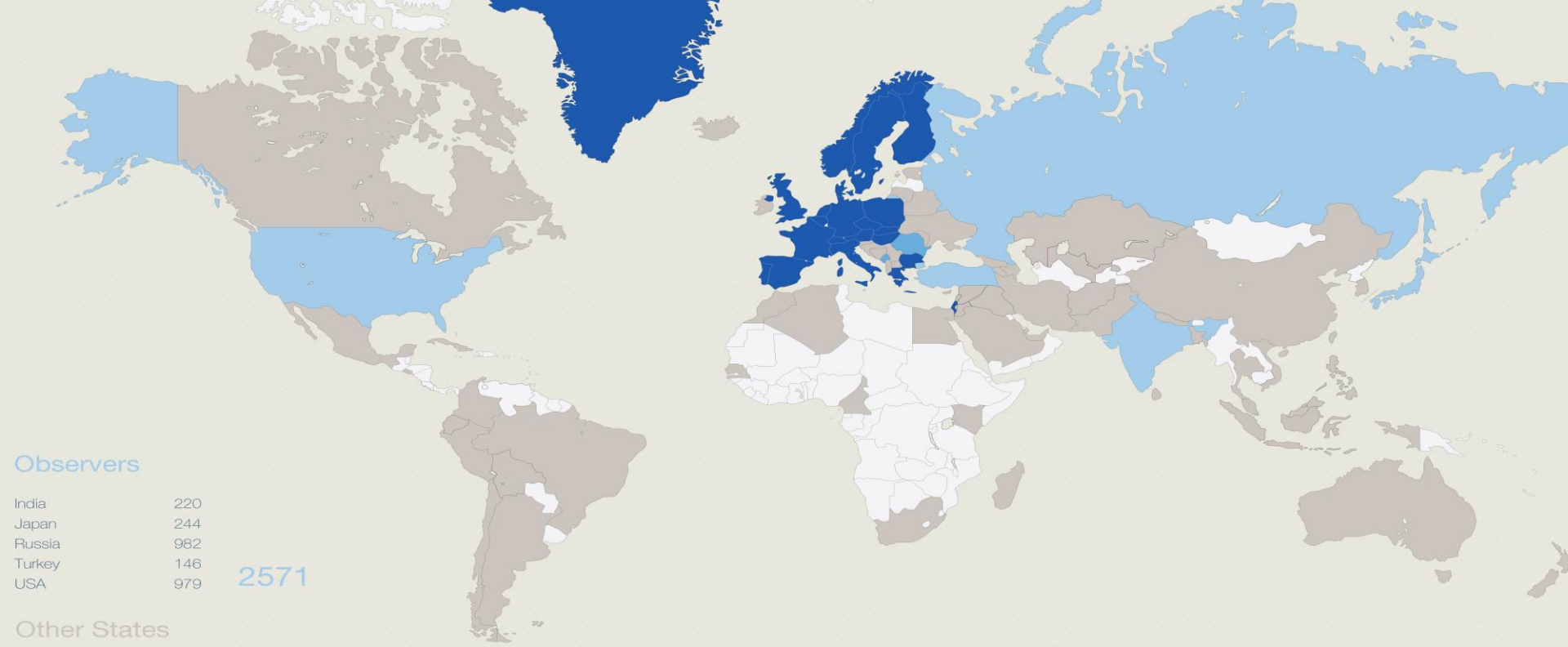
CERN Data Analytics Project

Improving CERN Accelerator Complex Operations with Data Analytics

Antonio Romero Marín

What is CERN

- CERN - European Laboratory for Particle Physics
- Founded in 1954 by 12 countries for fundamental physics research in a post-war Europe
  - "Science for Peace"



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

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

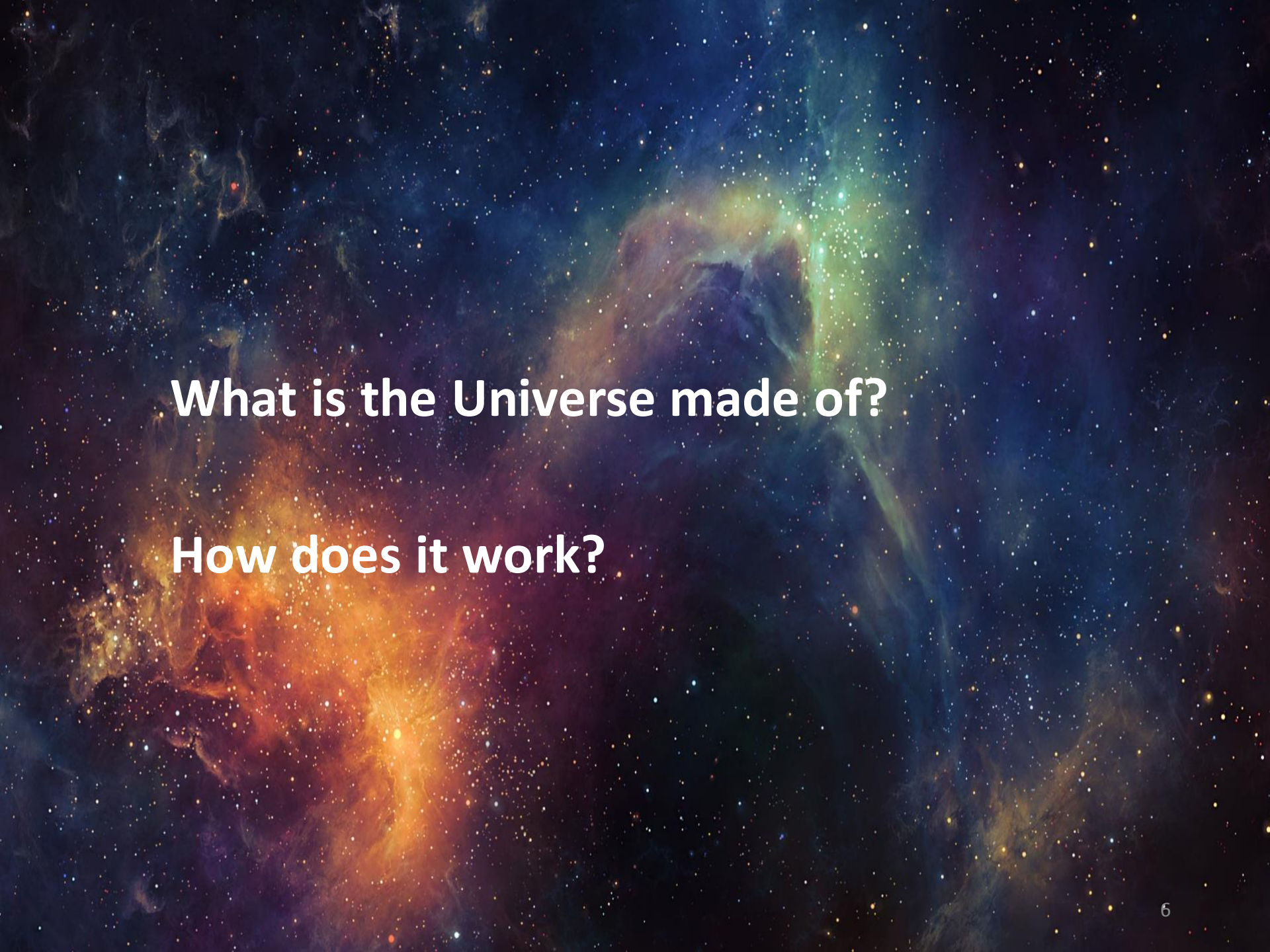
Serbia	41

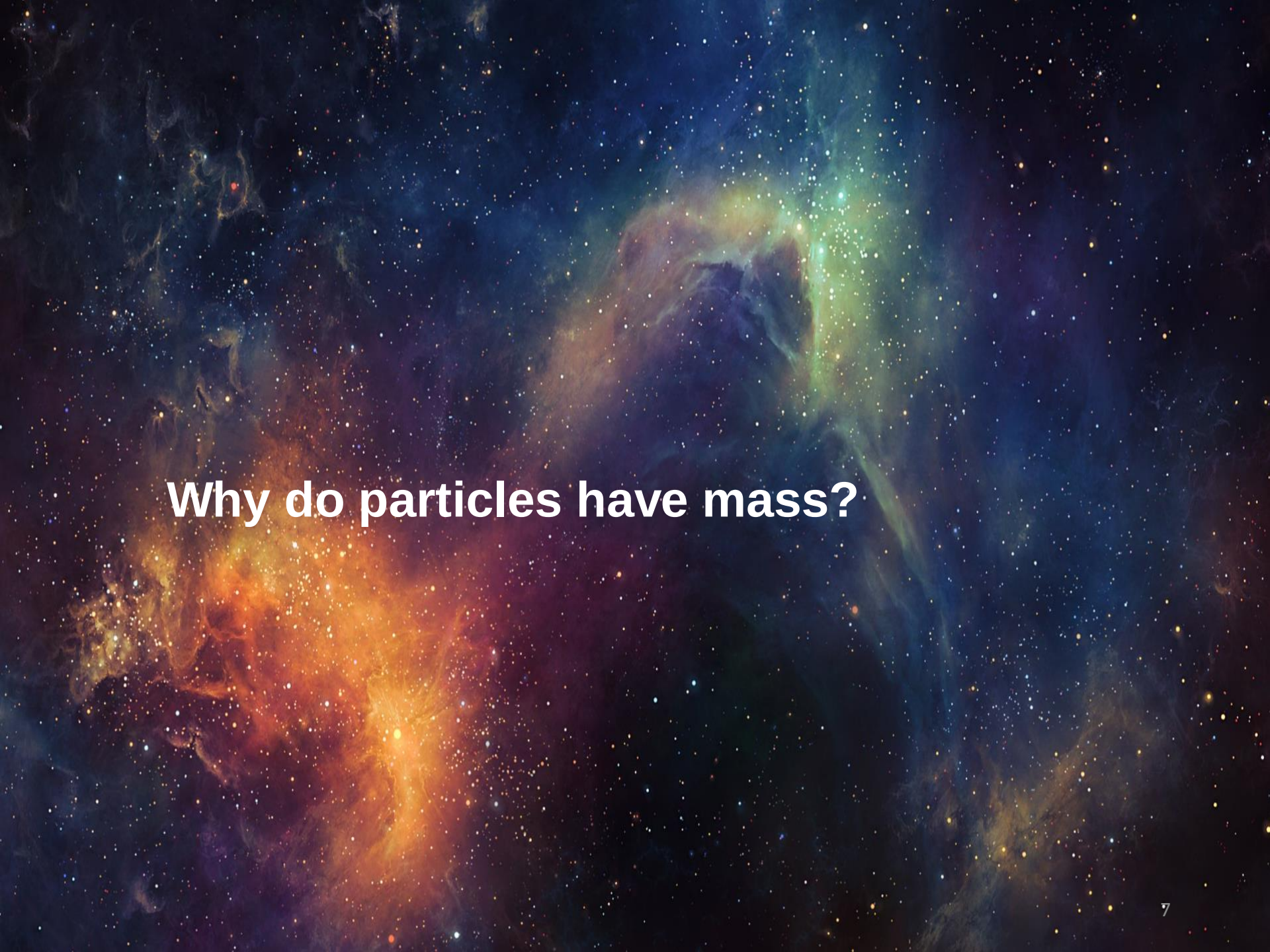
Distribution of All CERN Users by Nationality on 14 January 2014

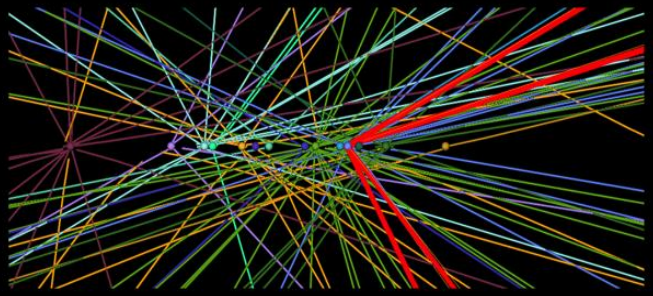
India at CERN

- Observer state
- Participating in experiments since the 1970s
- Collaborated in the construction
  - LHC accelerator
  - CMS and ALICE detectors
- Contributes to COMPASS, ISOLDE and nTOF experiments
- Operates two Tier-2 Centres for the LHC Computing Grid

CERN Mission


What is the Universe made of?
How does it work?


Why do particles have mass?

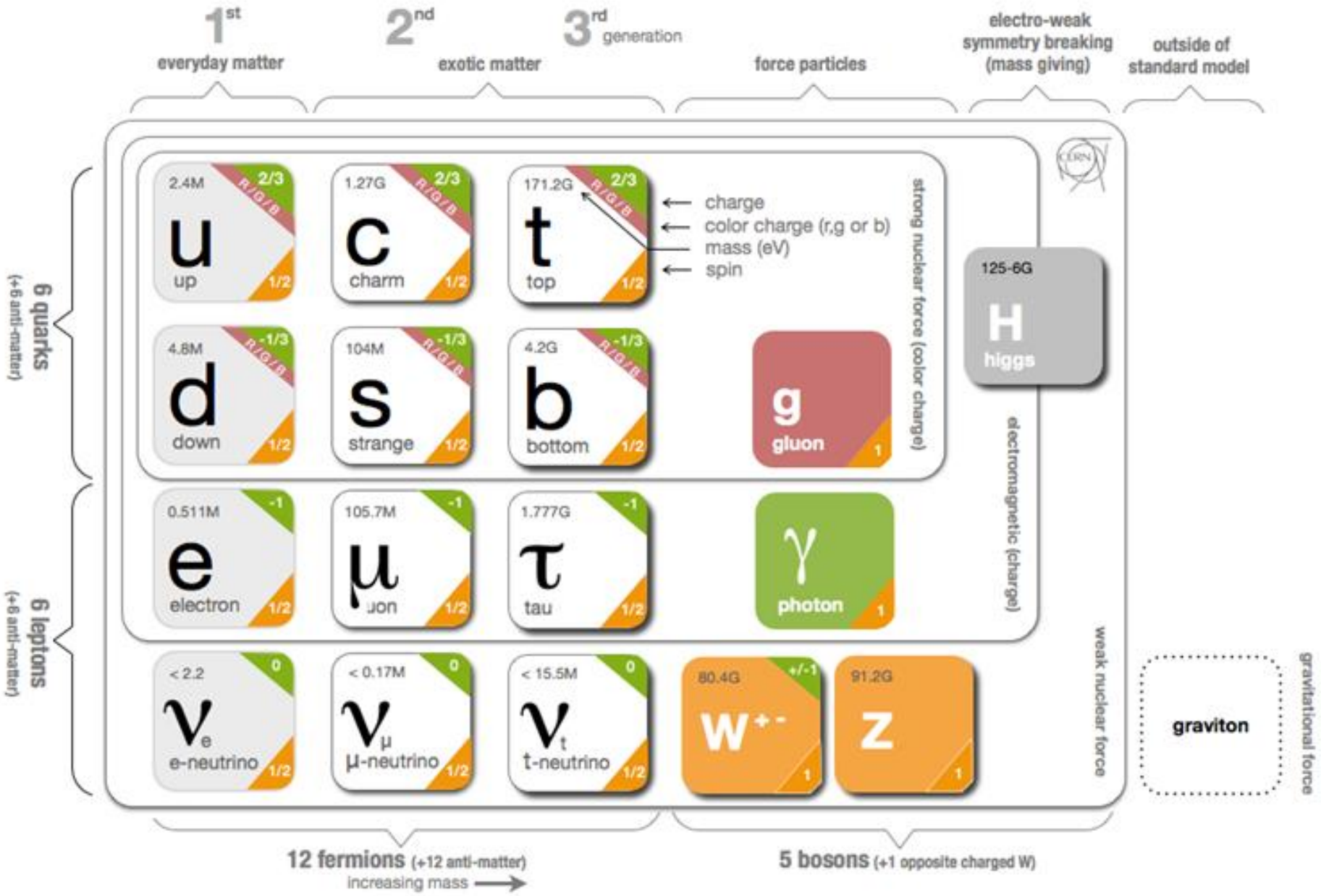


Higgs Boson Discovery

2012

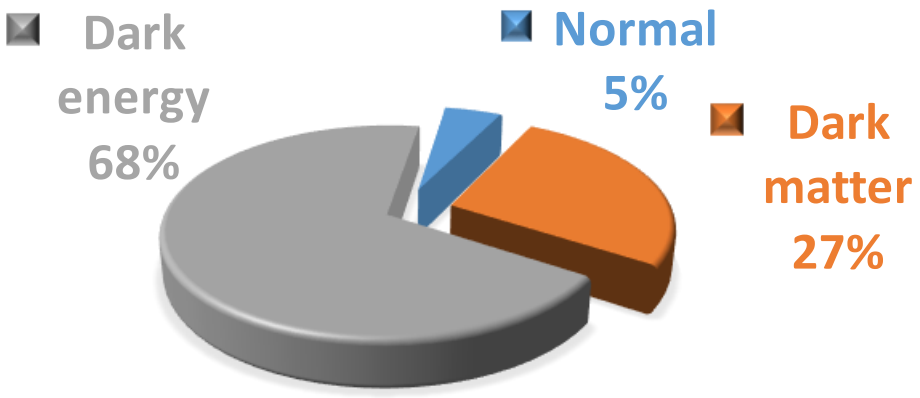
Higgs to 4μ candidate event

The Standard Model



Fundamental Research

- What is dark matter and dark energy?

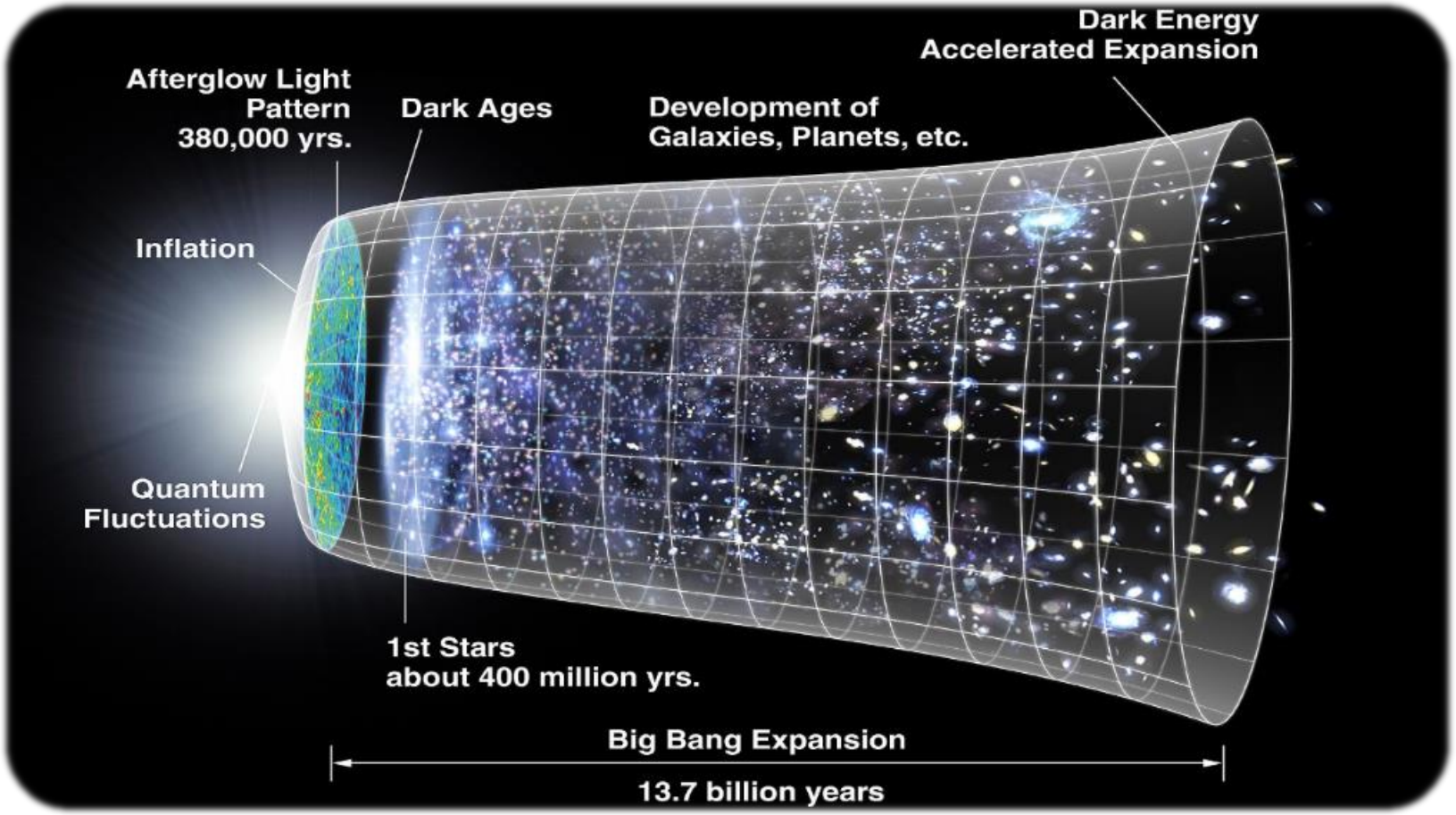


Composition of the Universe

- Why is there far more matter than antimatter in the universe?
  - Big Bang should have created equal amounts of matter and antimatter

Fundamental Research

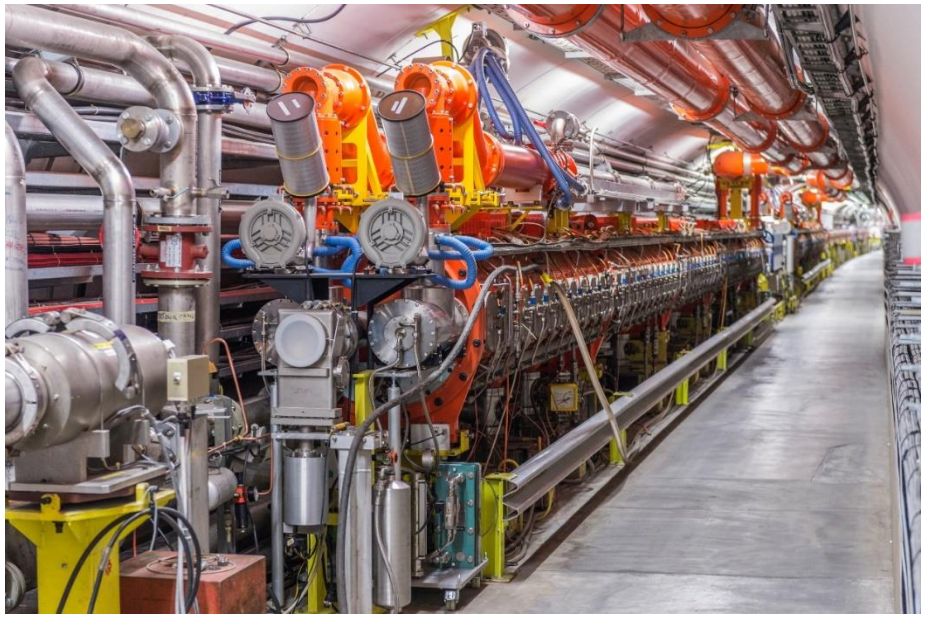
- What was the state of the matter in the very first moments of the Universe?



CERN Instruments

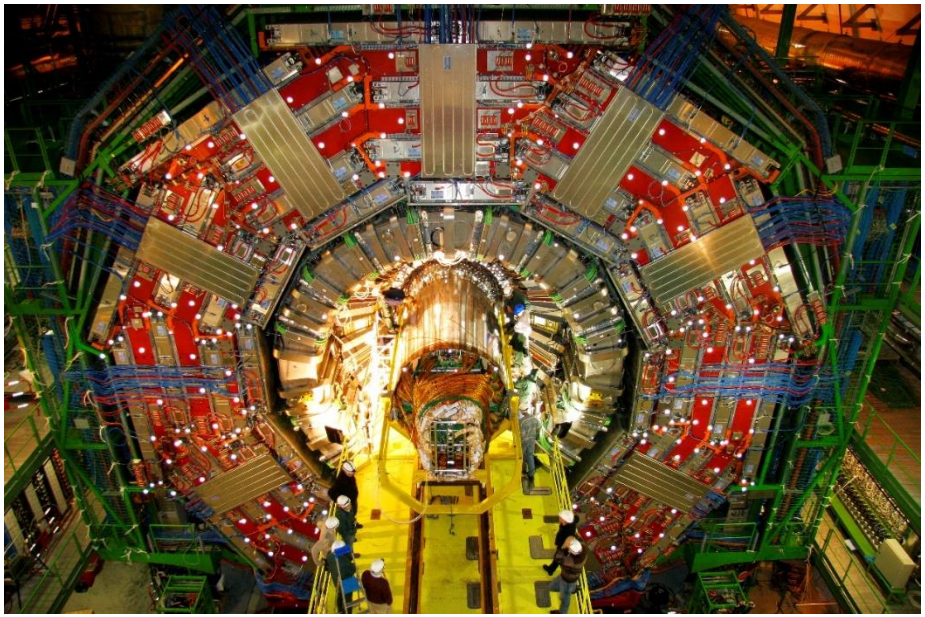
Accelerators

Boost particles to high energies and speed to collide



Detectors

Observe and record the results of these collisions



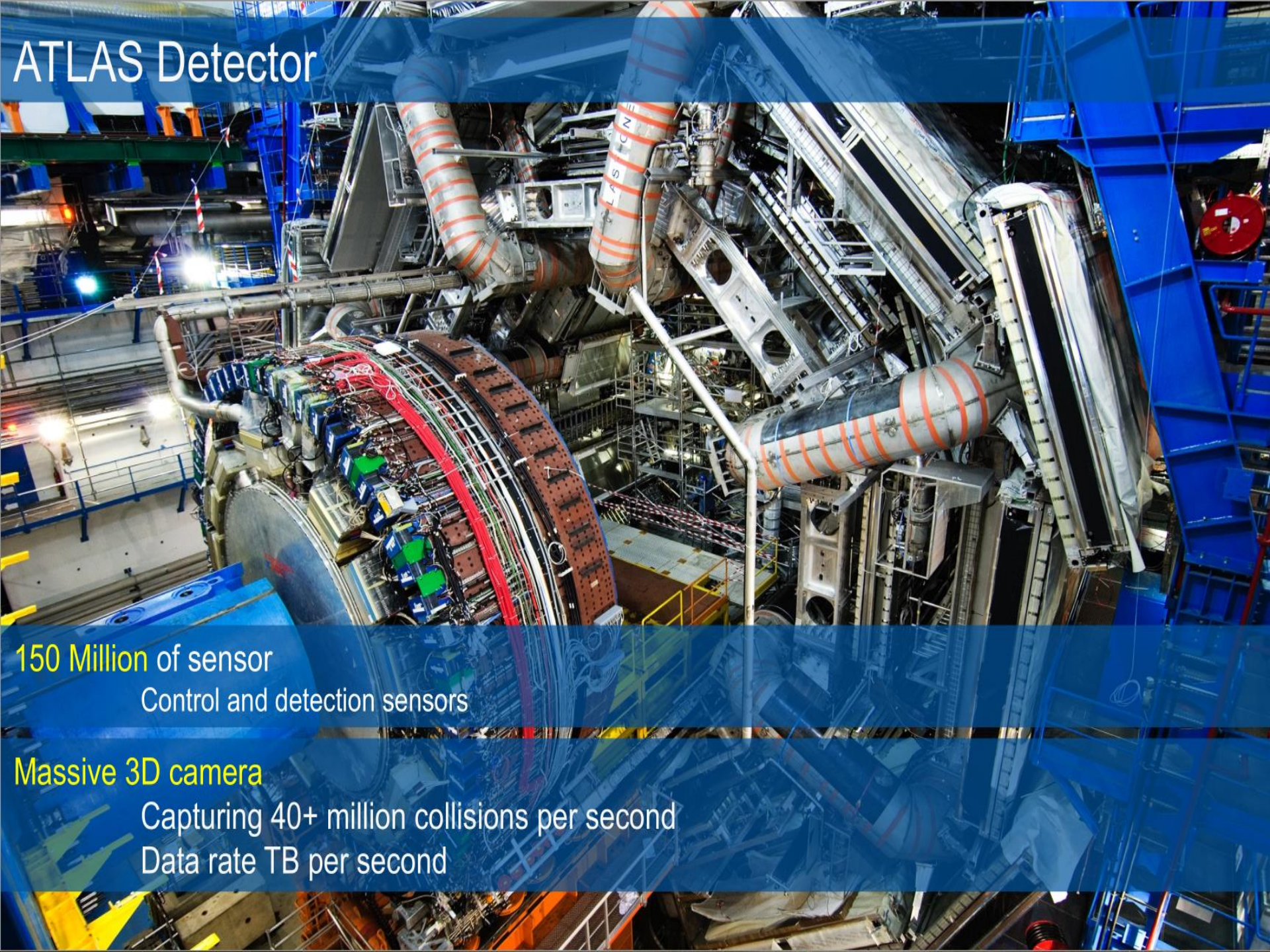
The Large Hadron Collider (LHC)

- Largest machine in the world
  - 27km, 6000+ superconducting magnets
- Fastest racetrack on Earth
  - Protons circulate 11245 times/s (99.9999991% the speed of light)
- 600 million collisions per second
  - Generating approximately one petabyte of data per second

The Large Hadron Collider (LHC)

- Emptiest place in the solar system
  - High vacuum inside the magnets
- One of the coldest places on Earth
  - Main magnets operate at a temperature of 1.9 K (-271.3°C)
- Hottest spot in the galaxy
  - During Lead ion collisions create temperatures 100000x hotter than the heart of the sun

ATLAS Detector



150 Million of sensor

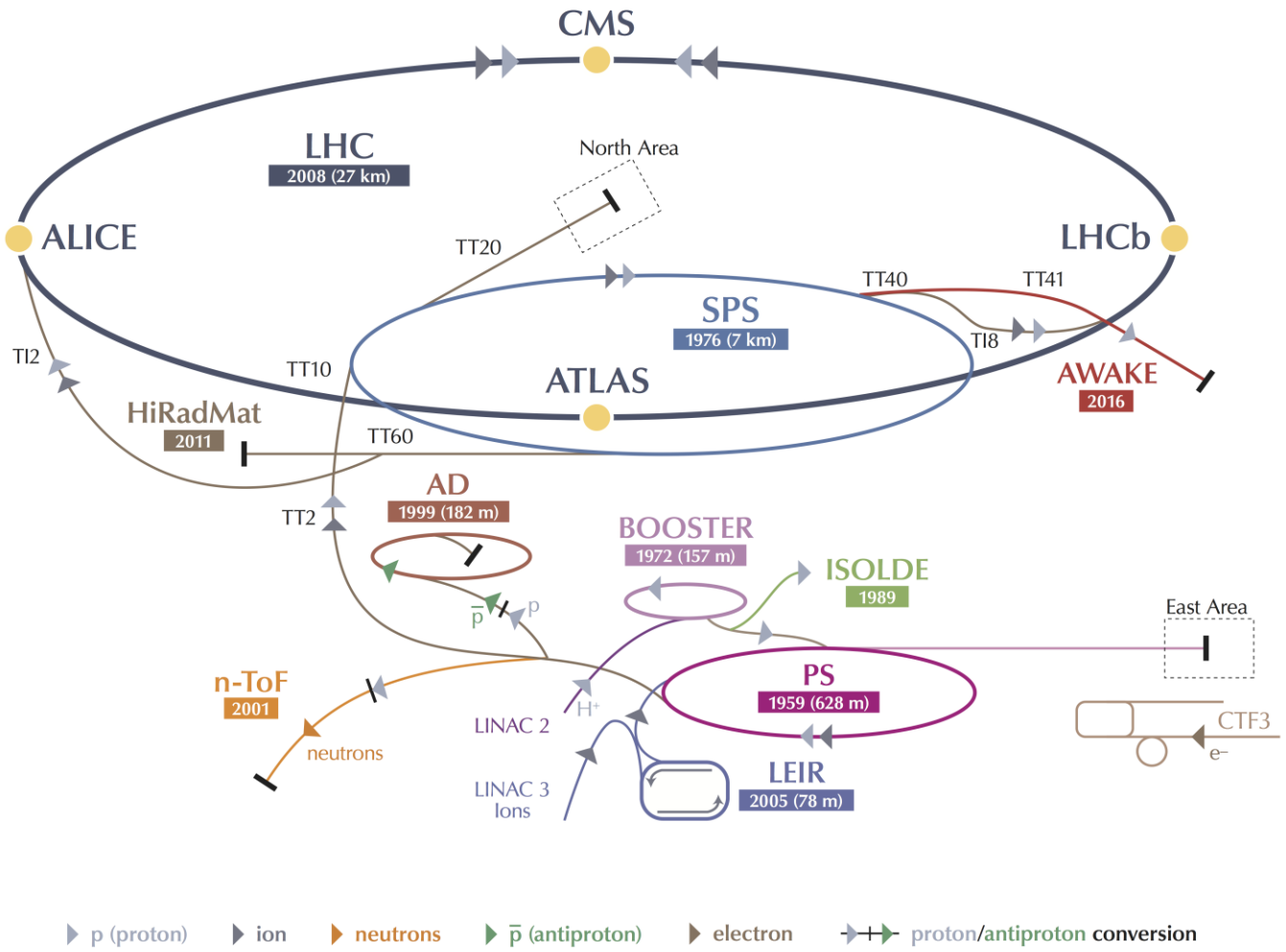
Control and detection sensors

Massive 3D camera

Capturing 40+ million collisions per second

Data rate TB per second

The CERN Accelerator Complex

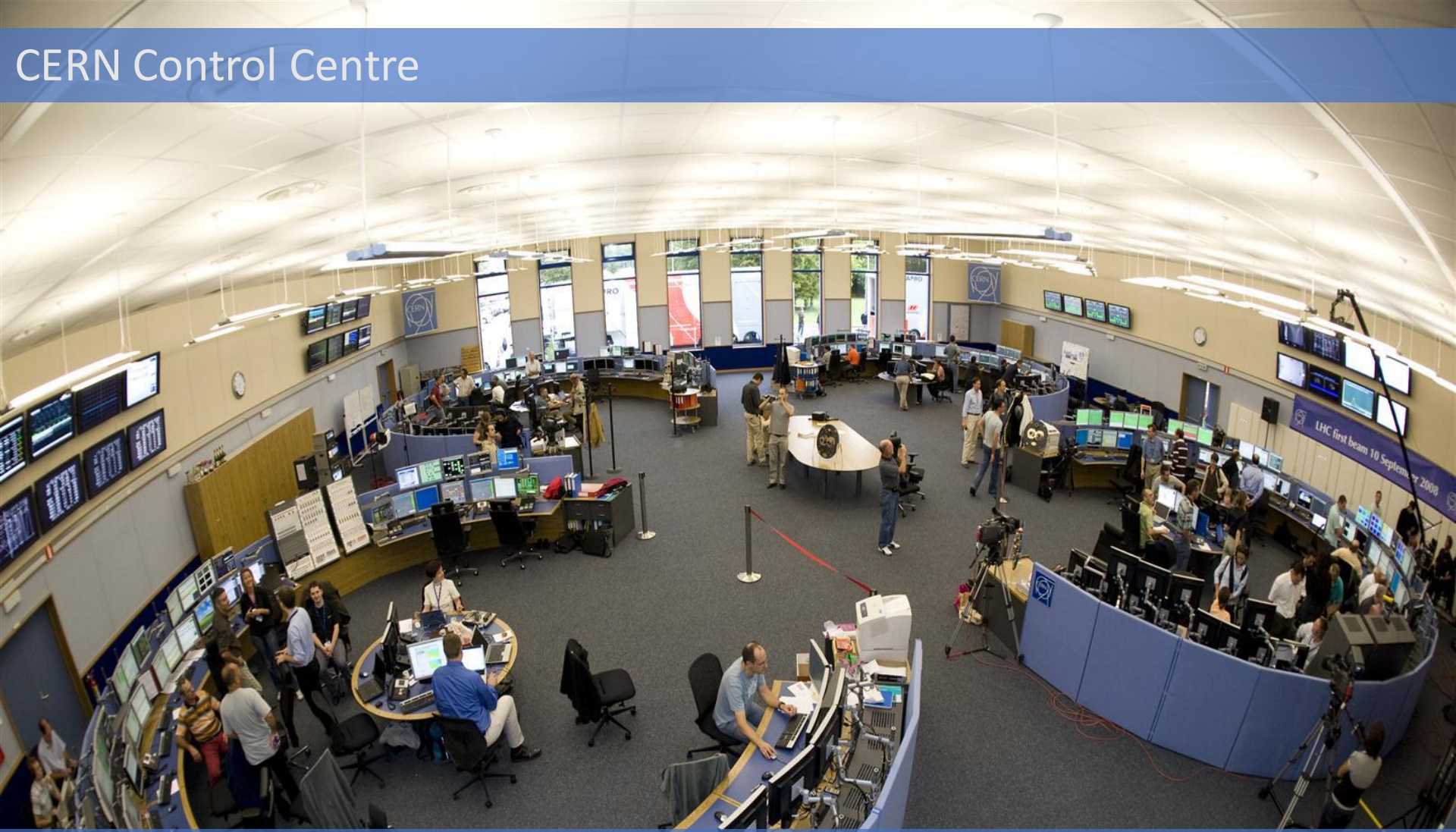


LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron

AD Antiproton Decelerator CTF3 Clic Test Facility AWAKE Advanced WAKefield Experiment ISOLDE Isotope Separator OnLine DEvice

LEIR Low Energy Ion Ring LINAC LINear ACcelerator n-ToF Neutrons Time Of Flight HiRadMat High-Radiation to Materials

CERN Control Centre



CERN Accelerator Complex is unique installation

Therefore, we have to face unique challenges

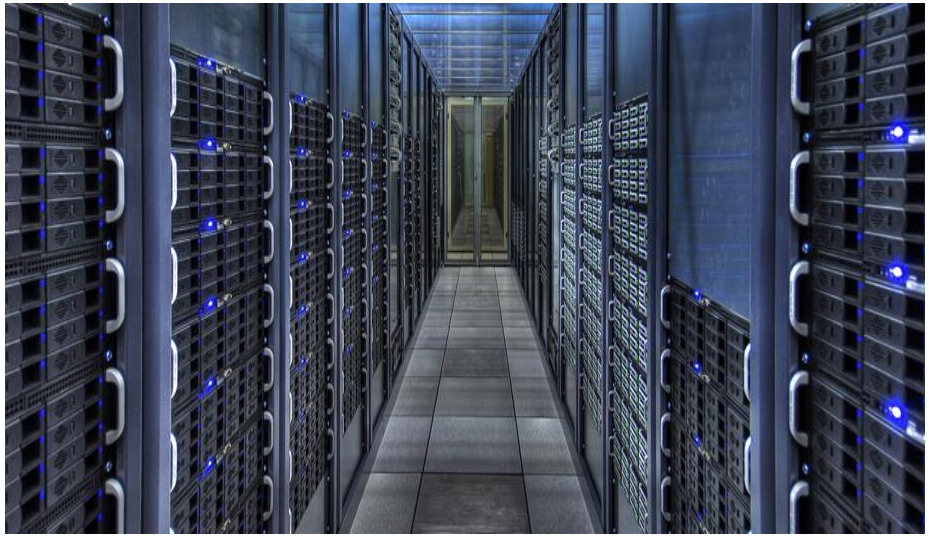
Control and operations

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

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

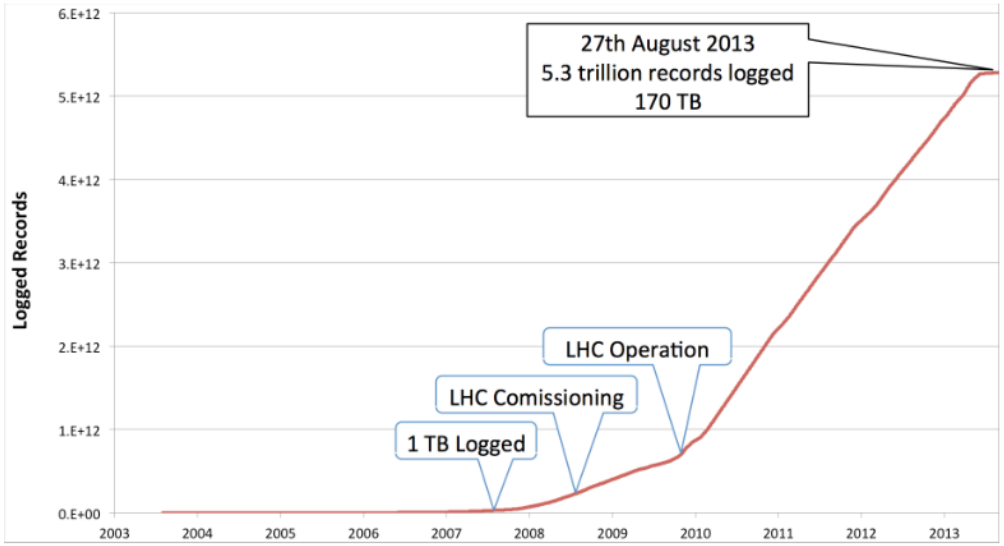
CERN is a extreme data environment

- Physics Data
  - Over 30 petabytes per year
- Control and operations
  - Million of sensors, signals
  - Large number of control devices
  - Equipment
- Monitoring and logging
- Supporting IT infrastructure
  - Databases
  - Network
  - Services



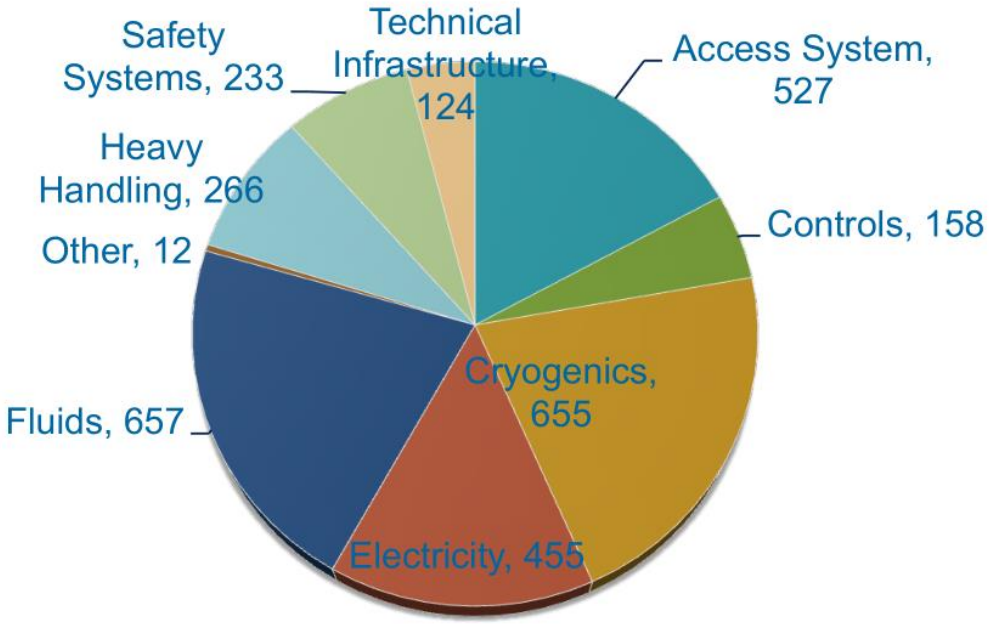
CERN Data Investment

- CERN has great monitoring and logging systems
  - Large amount of data has been stored over years
- Accelerator Logging Service around 275 GB/day
  - Storing more than 50 TB / year
  - Data aquisition
    - CERN accelerator complex
    - related subsystems
    - experiments
  - Around 1 million signals



Data Analytics Challenges

- Some faults cannot be avoid
- Decrease the availability for running physics
- Corrective interventions needed

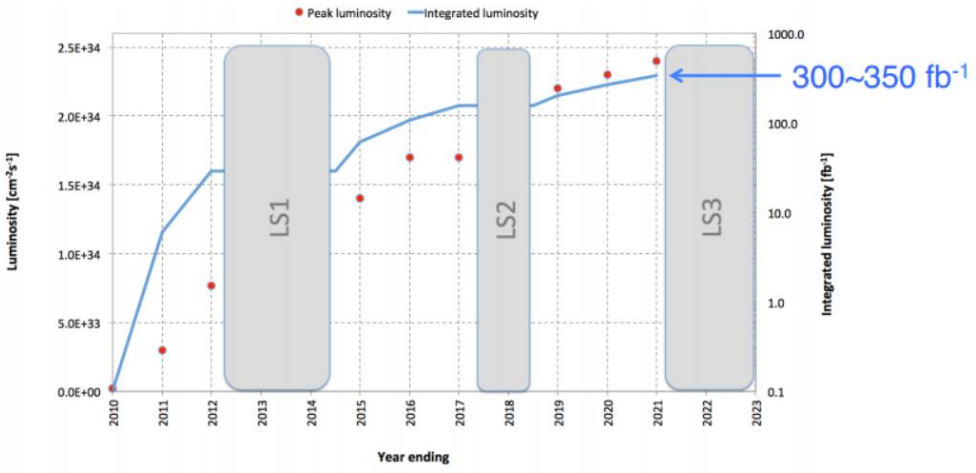


A look into the Future

- LHC upgrades will further increase luminosity
  - Computing resources needs will be higher
  - Data generated will increase drastically

Table 17: HLT Rate Evolution

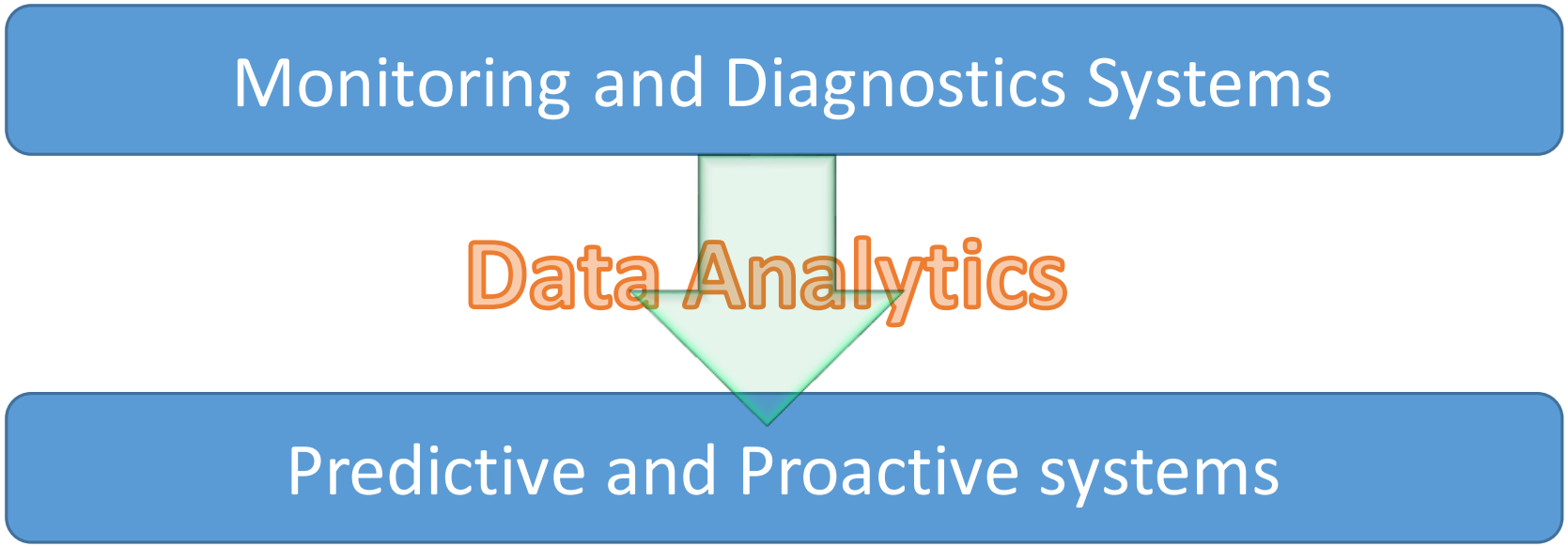
Hz	ALICE	ATLAS	CMS	LHCb
2012	400 Hz 330 MB/s (p-p) 540 MB/s (p-Pb)	550 Hz 440 MB/s	460+360 Hz 328 MB/S	5000 Hz 300 MB/s
2015	500 Hz 525 MB/s (p-p) 810 MB/s (p-Pb) 3750 MB/s (Pb-Pb)	1000 Hz 800-1000 MB/s	1000 Hz 600 MB/S	10000 Hz 750 MB/s



- Next accelerators
  - Future Circular Collider (80-100 km)

FCC hh ee he

The objective – Improve our systems



Data Analytics Project

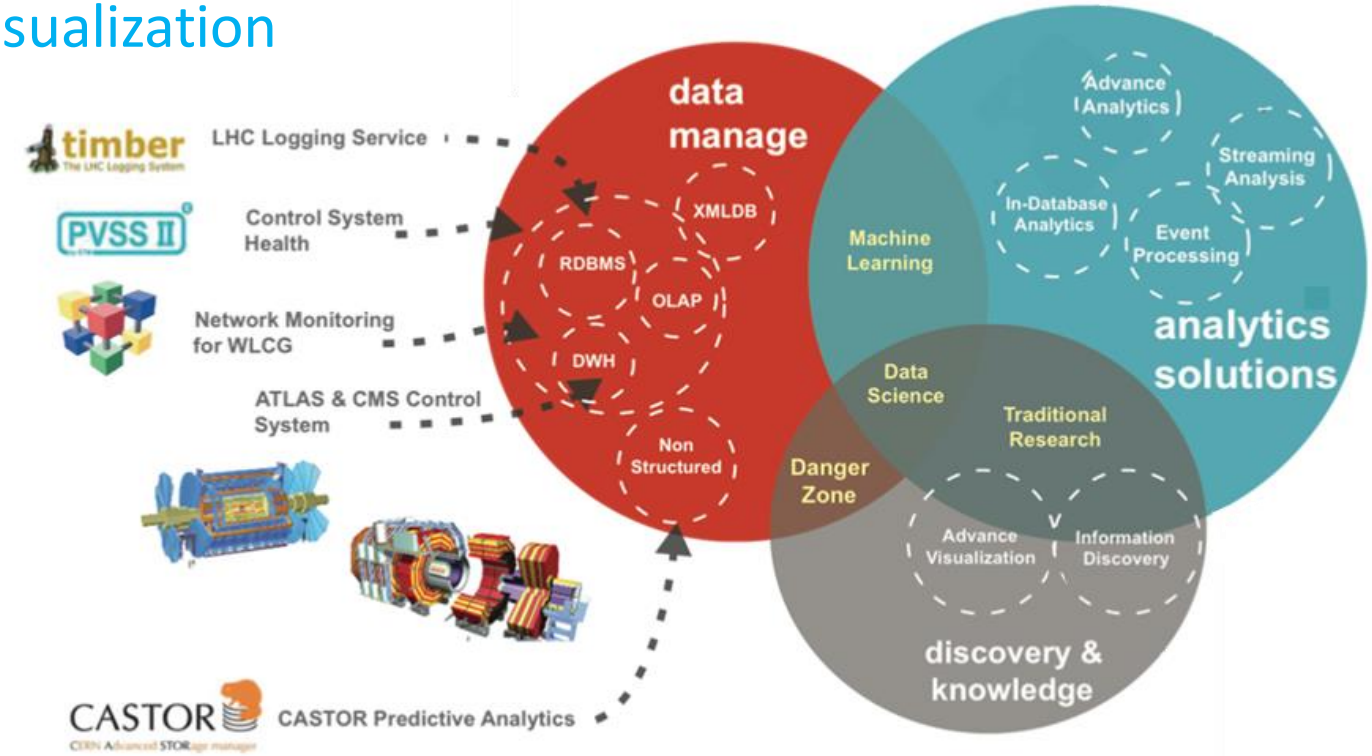
- Optimize our systems
  - Reducing and predicting faults and corrective interventions
  - Increase the availability and operations efficiency
- Profit from CERN data investment by using data analytics
  - Extract knowledge
  - Discover useful information
  - Suggest conclusions
  - Support decision making
- Control and Monitoring Systems
  - Proactive
  - Predictive
  - Intelligent

Data analytic challenges at CERN

- Very dynamic and heterogeneous environment
  - Large number projects with different needs
  - Technologies
  - Data sources
- Large amount of data
  - Raw data, structured and unstructured data
- Education and Training
  - Users know their systems and data
  - Help them to use data analytics

Data analytic challenges at CERN

- Analytics as a service
  - Analytics platform
  - Multiple sources of data
  - Visualization
  - Data processing
  - Easy access



CERN openlab

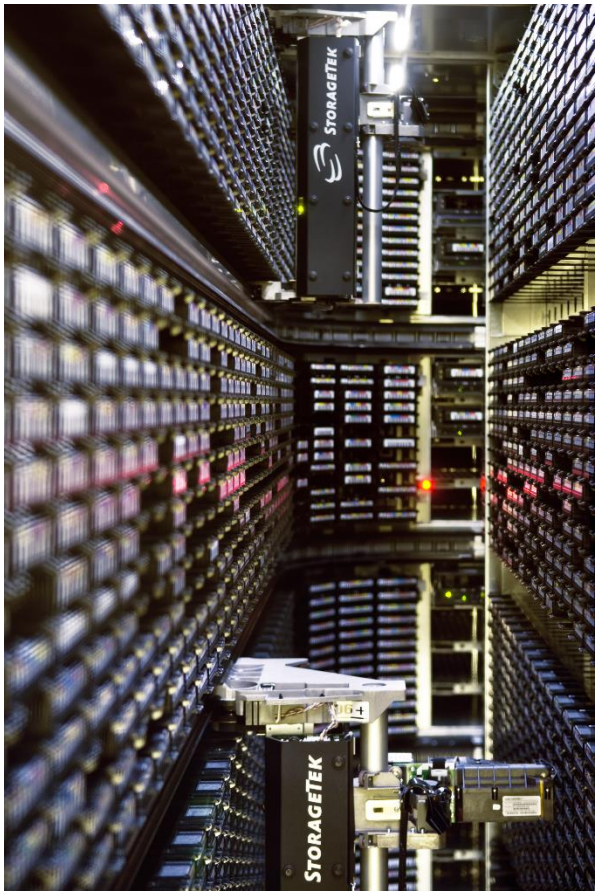
- Public-private partnership between CERN and leading ICT companies
- Accelerate cutting-edge solutions to be used by the worldwide LHC community
- Designed to create and disseminate knowledge
  - Publication of reports and articles
  - Workshops or seminars
  - CERN openlab Student Programme

Data Analytics Use Cases

Use Case: CERN Advance Storage Manager (CASTOR)

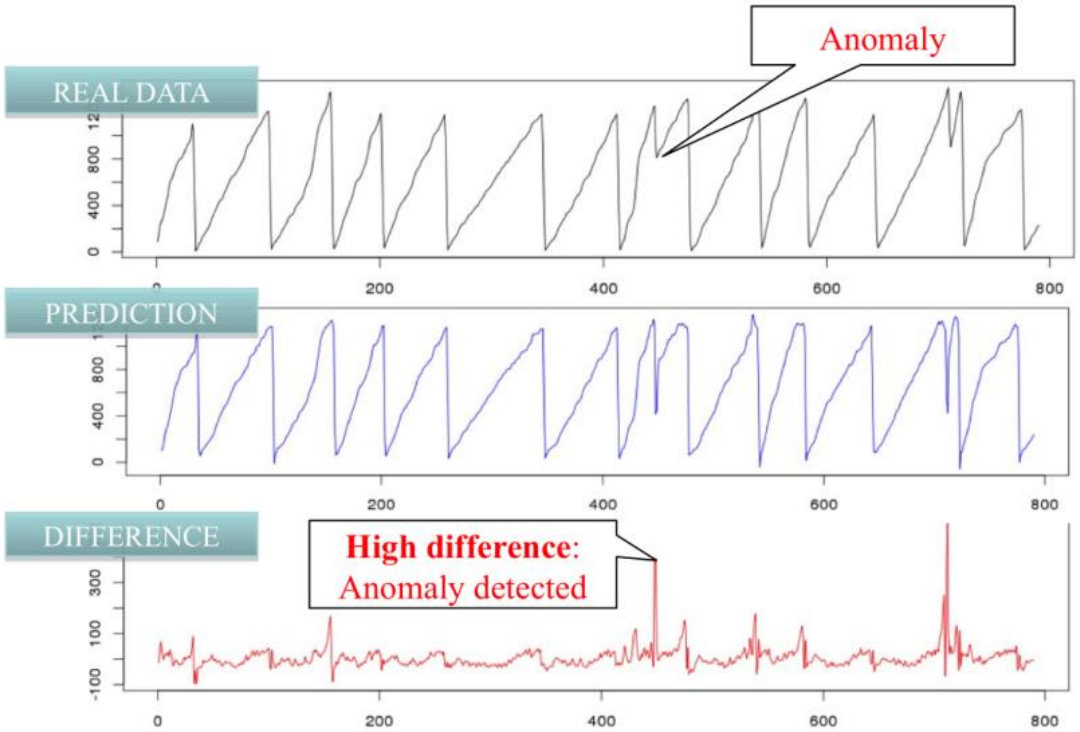
- Mass Storage Solution for managing physics data files
  - 12k disks, 30k tapes
  - 100 PB on tape, 50 PB on disk
  - +300 M files

CASTOR
CERN Advanced STORage manager



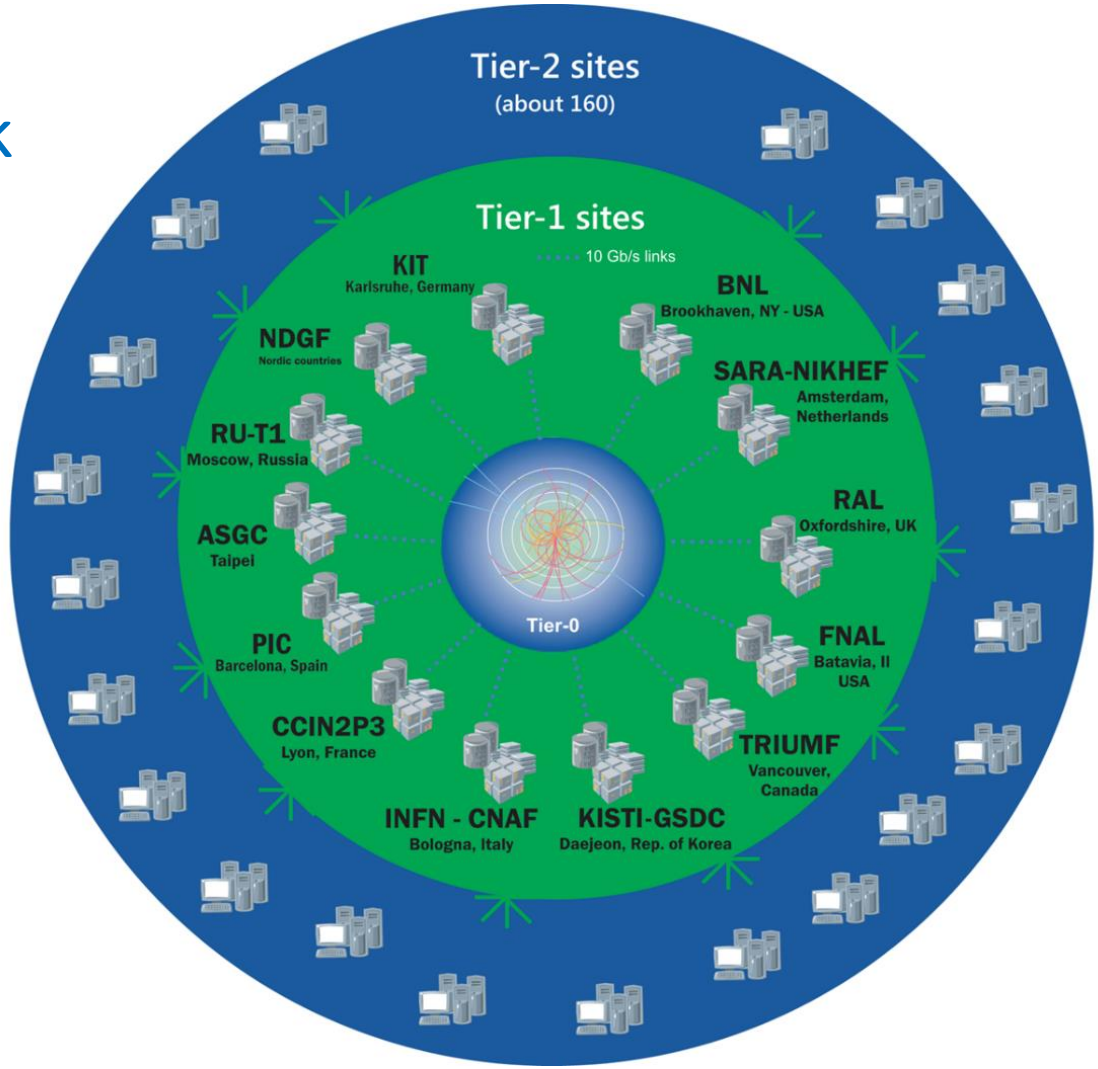
Use Case: CERN Advance Storage Manager (CASTOR)

- Optimization
  - Performance
  - Cause of errors
- Predictive analytics
  - Anomaly detection
  - Early warning systems



Use Case: Intelligent Data Placement for CMS

- Collaborative network of data centers to transmit, store, process and analyze LHC data
- Over 8000 physicists



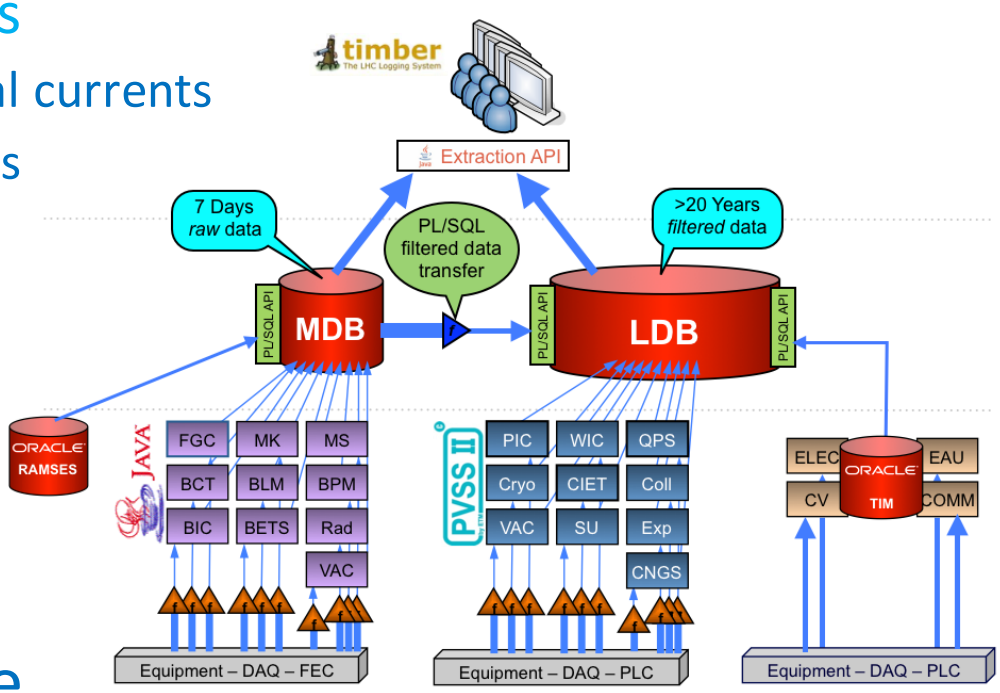
WLCG
Worldwide LHC Computing Grid

Use Case: Intelligent Data Placement for CMS

- CMS Data Popularity
  - Monitors along time the usage of data accessed by users
- Resources optimization
  - Minimize number of replicas
  - Remove Obsolete
  - Job time in data access
  - Job time in data analysis
- Resources Prediction

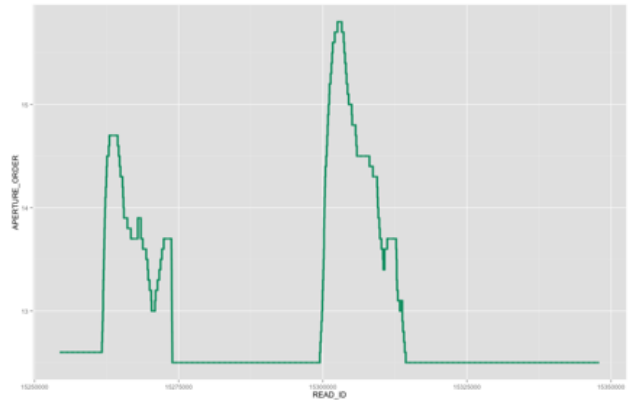
Use Case: Operation and Control Systems

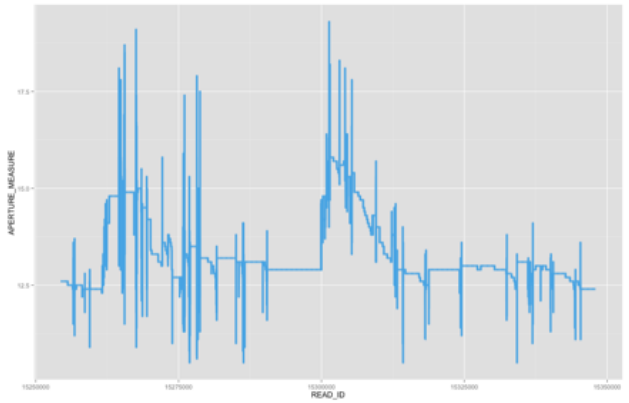
- LHC Accelerator Logging Service
  - Around 1 million signals
    - Temperatures, electrical currents
    - Magnetic field strengths
    - Vacuum pressures, etc.
- Control system Health
  - Gas Breakdown
  - Vacuum
  - Machine Protection
- Predictive maintenance

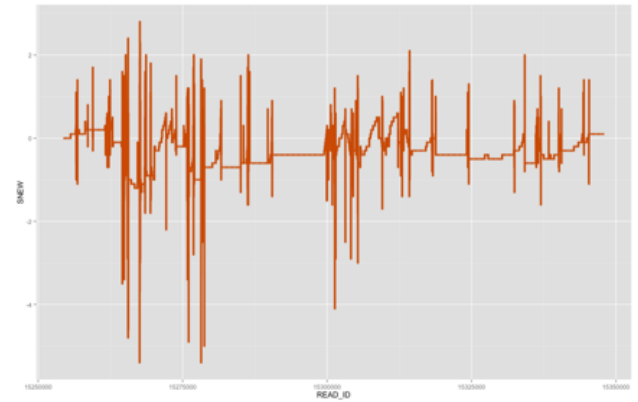


Use Case: Cryogenics Faulty Valves Detection

- What is the objective?
  - Predict faulty valves before they actually fail
- How?
  - Valve receive an aperture order value (**aperture order**)
  - Effective aperture realized by the valve (**aperture measured**)
  - Analyzing the difference between both (**S** = **aperture order - aperture measured**)







Use Case: Cryogenics Faulty Valves Detection

- Signals used
  - S = aperture order - aperture measured
- Features extractions based on S
  - Variance
  - Percentile 99.9
  - Rope distance
  - Noise Band
- Three different status
  - Faulty
  - Not faulty
  - Unknown
- **Predictive model**
  - SVM - Support Vector Machine

Additional comments

- We used “supervised machine learning”
  - Finely tuned predictor function
  - Training with dataset with the “right answers”
- But it is important to process the data in real-time
  - Data streaming and Complex Event Processing
  - Early warning systems
  - Real-time data visualization and discovery
- New platforms aim to facilitate the use
  - Quick development and prototype of predictive models
  - Multiple languages (Scala, R, Python, Java, etc…)
  - Complete stack
  - Cloud based

www.cern.ch