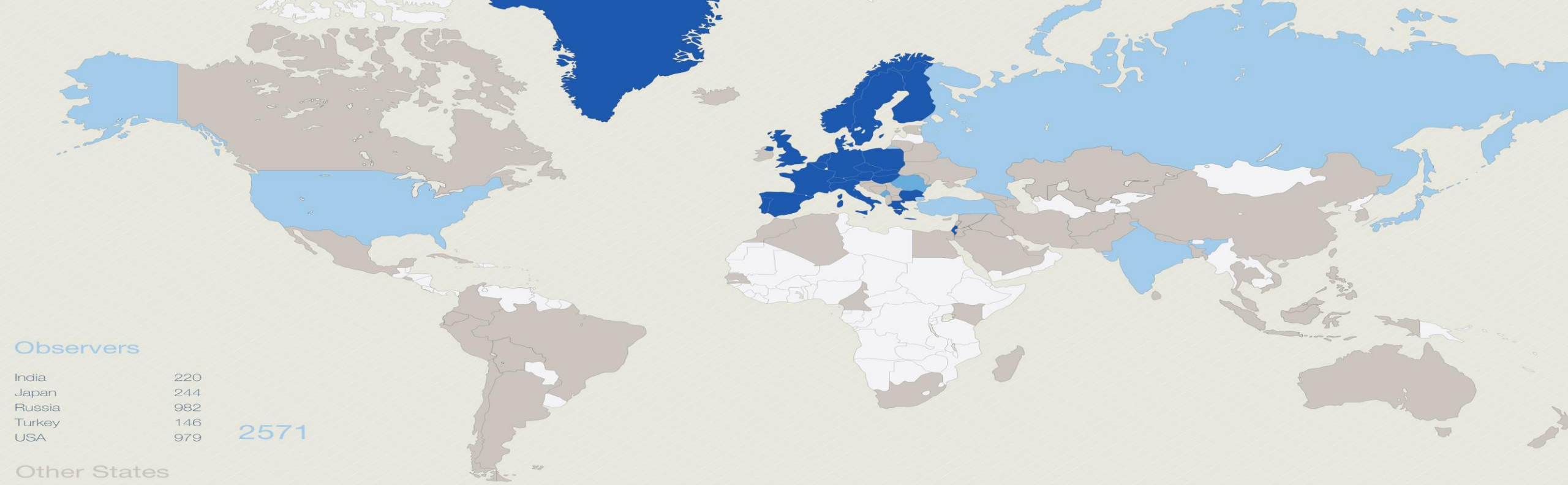
Big Data Analytics
for the Exploitation of the CERN Accelerator Complex

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

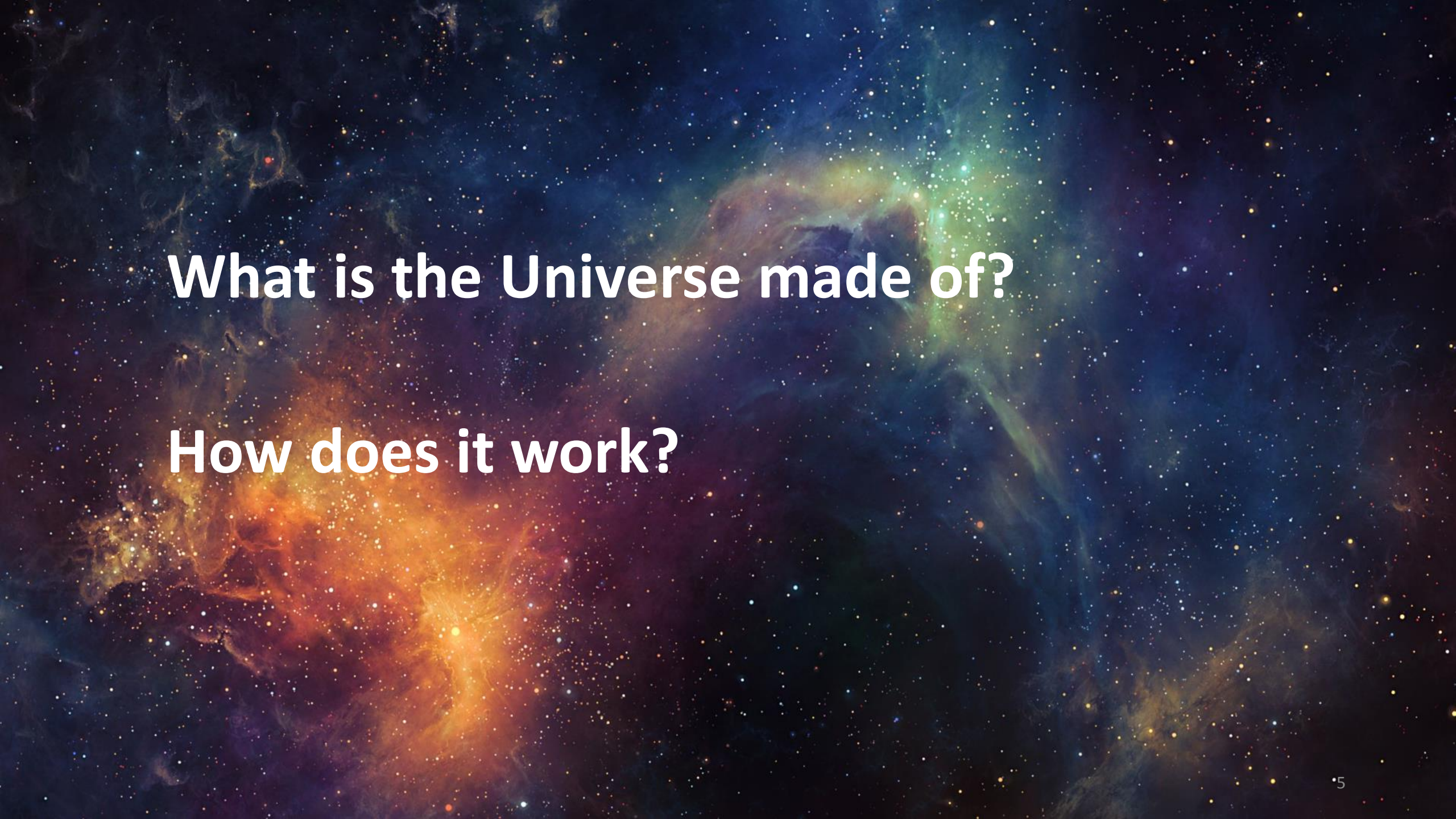
Romania	118

Associate Members in the Pre-stage to Membership

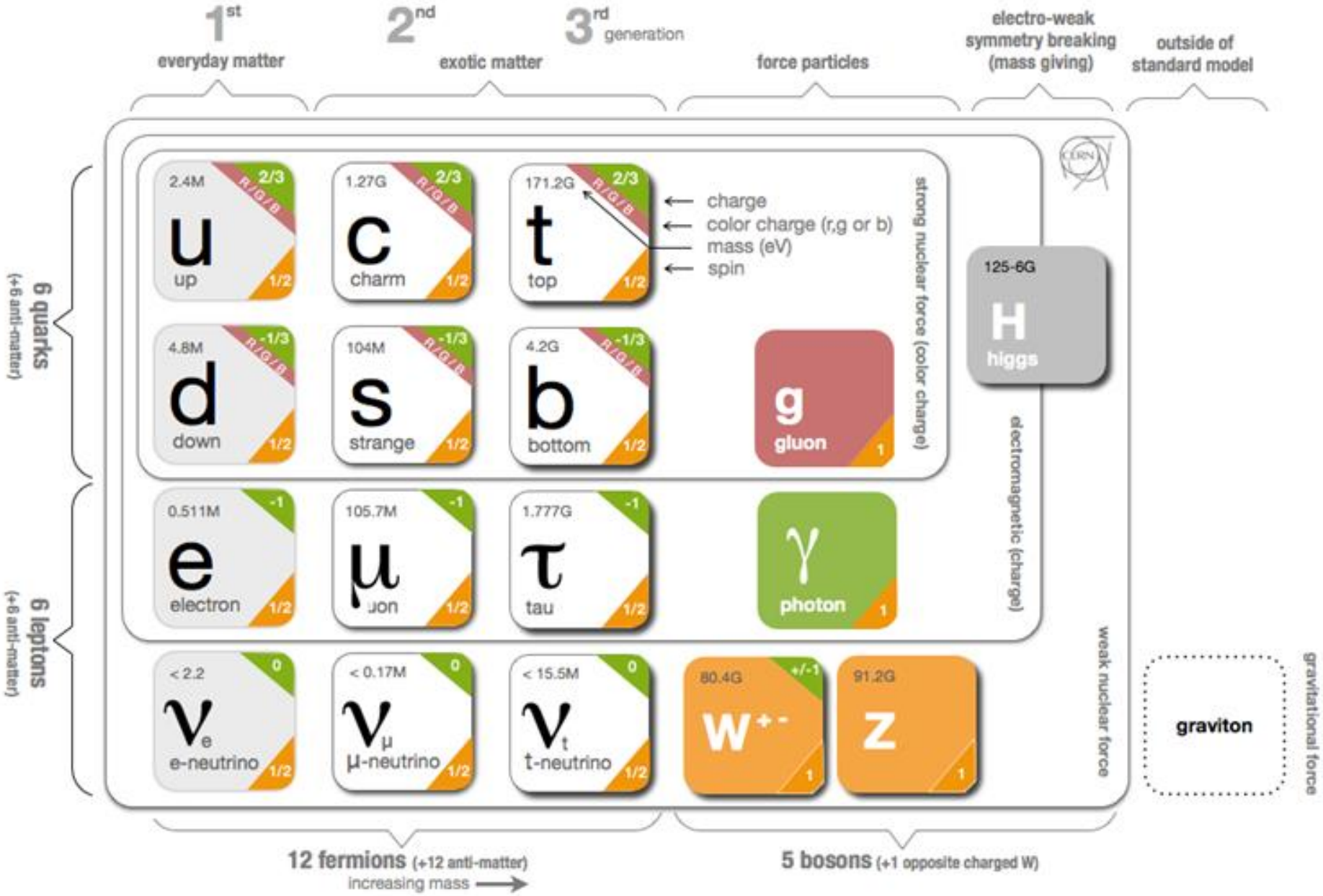
Serbia	41

Distribution of All CERN Users by Nationality on 14 January 2014

CERN Mission

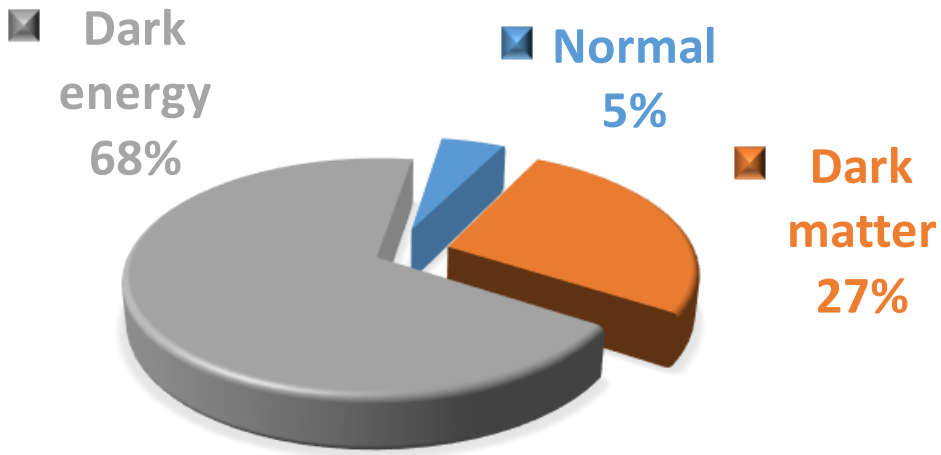

What is the Universe made of?
How does it work?

The Standard Model



Fundamental Research

- What is dark matter and dark energy?

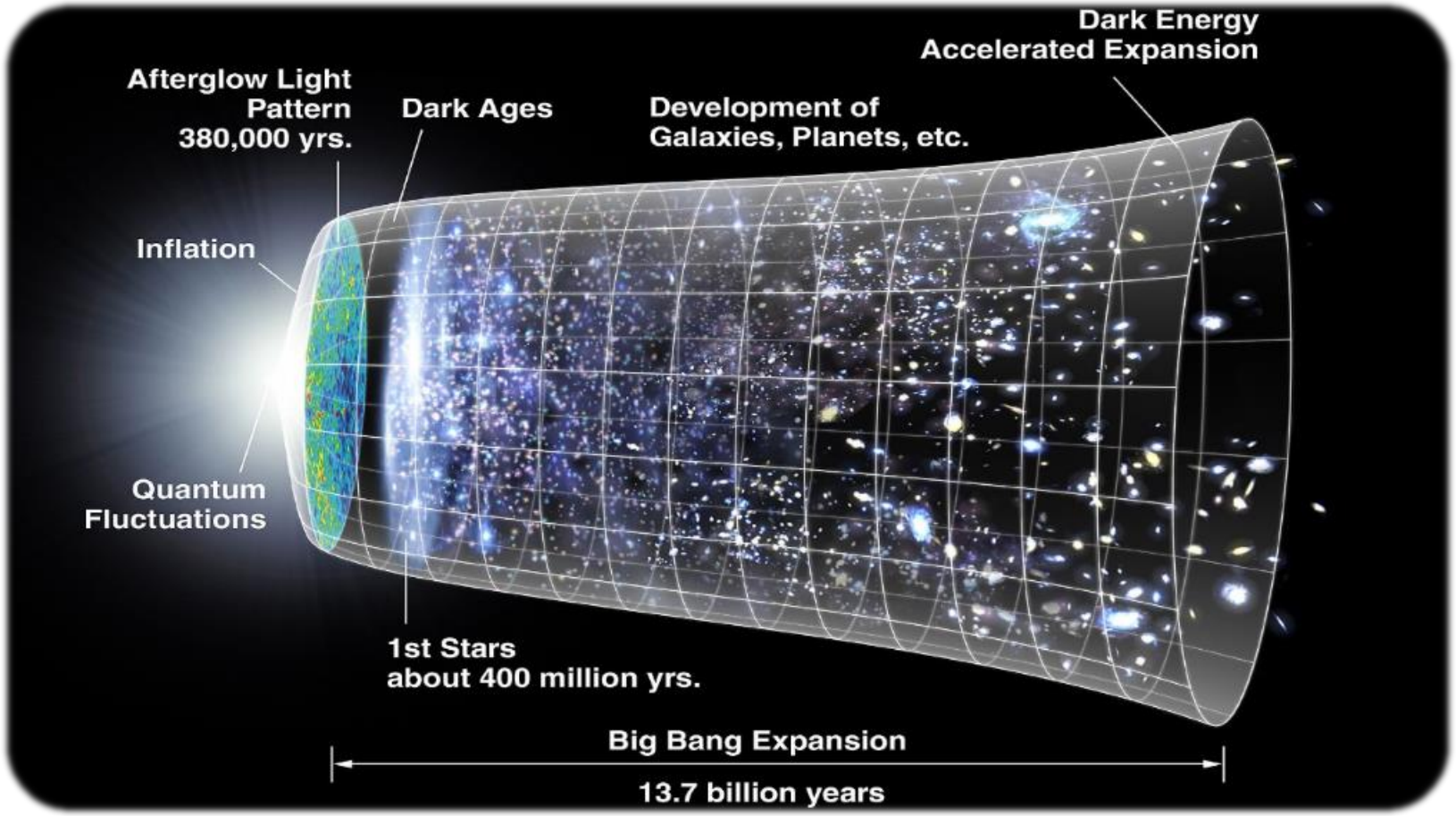


Composition of the Universe

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

Fundamental Research

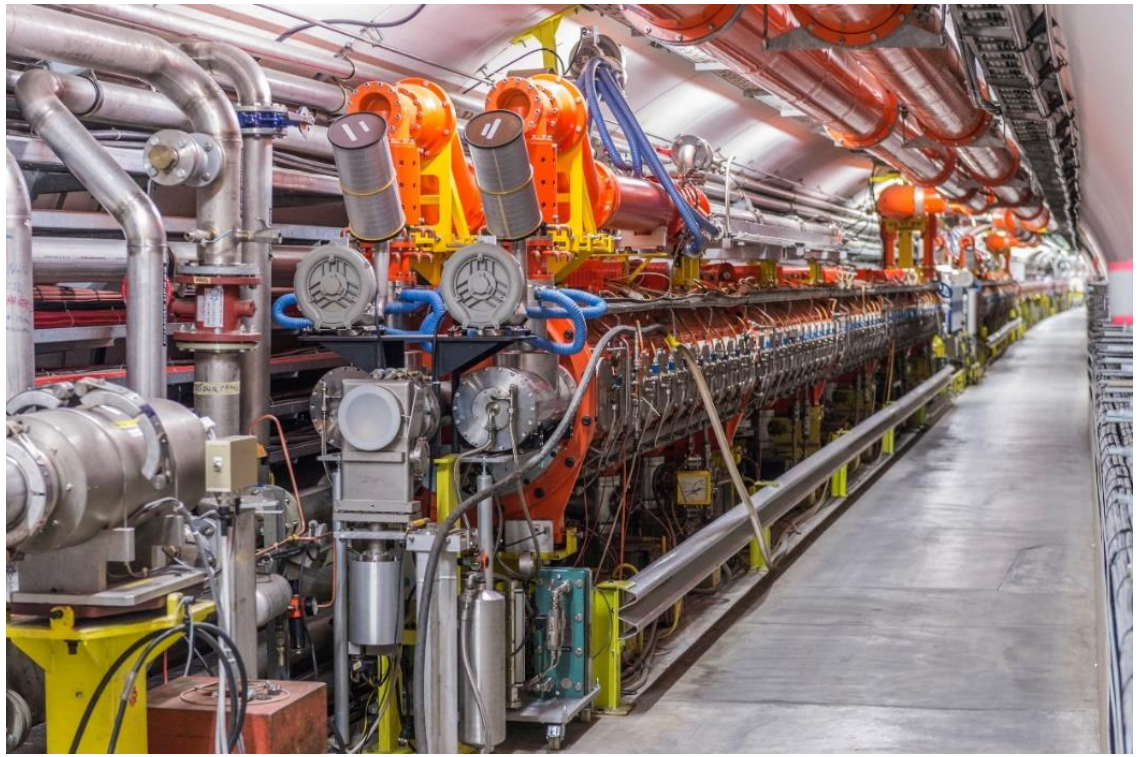
- What was the state of the matter in the very first moments of the Universe?



CERN Instruments

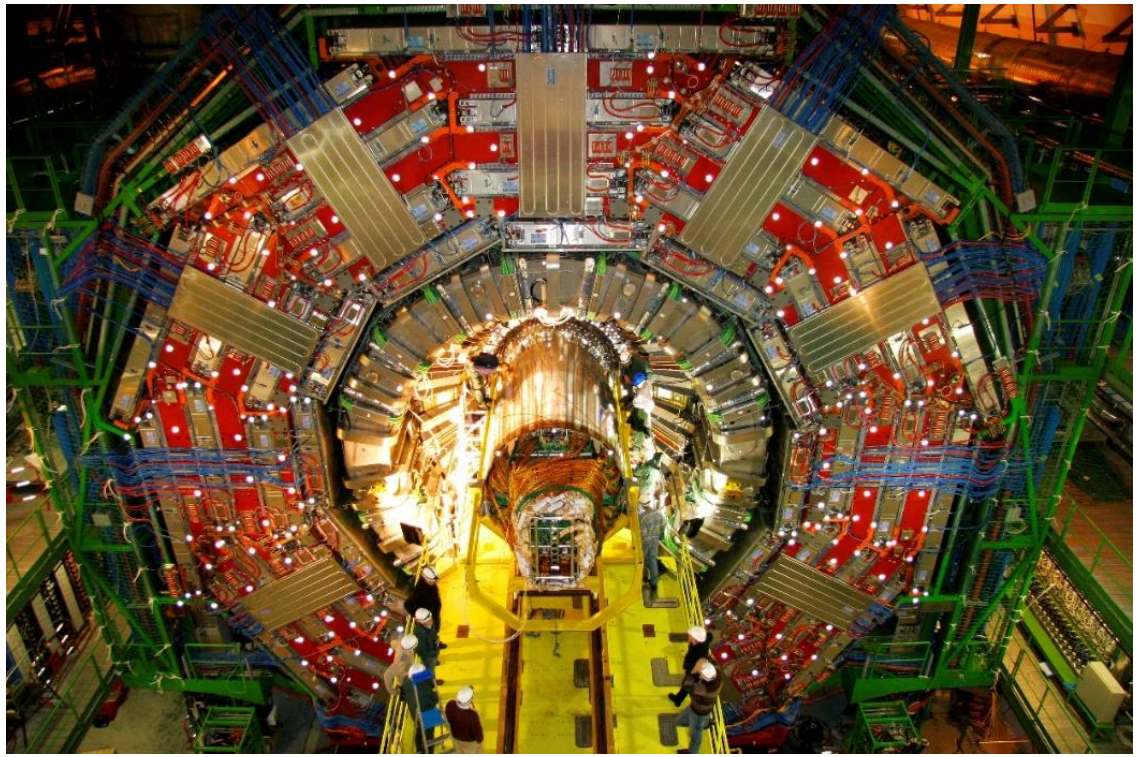
Accelerators

Boost particles to high energies and speed to collide



Detectors

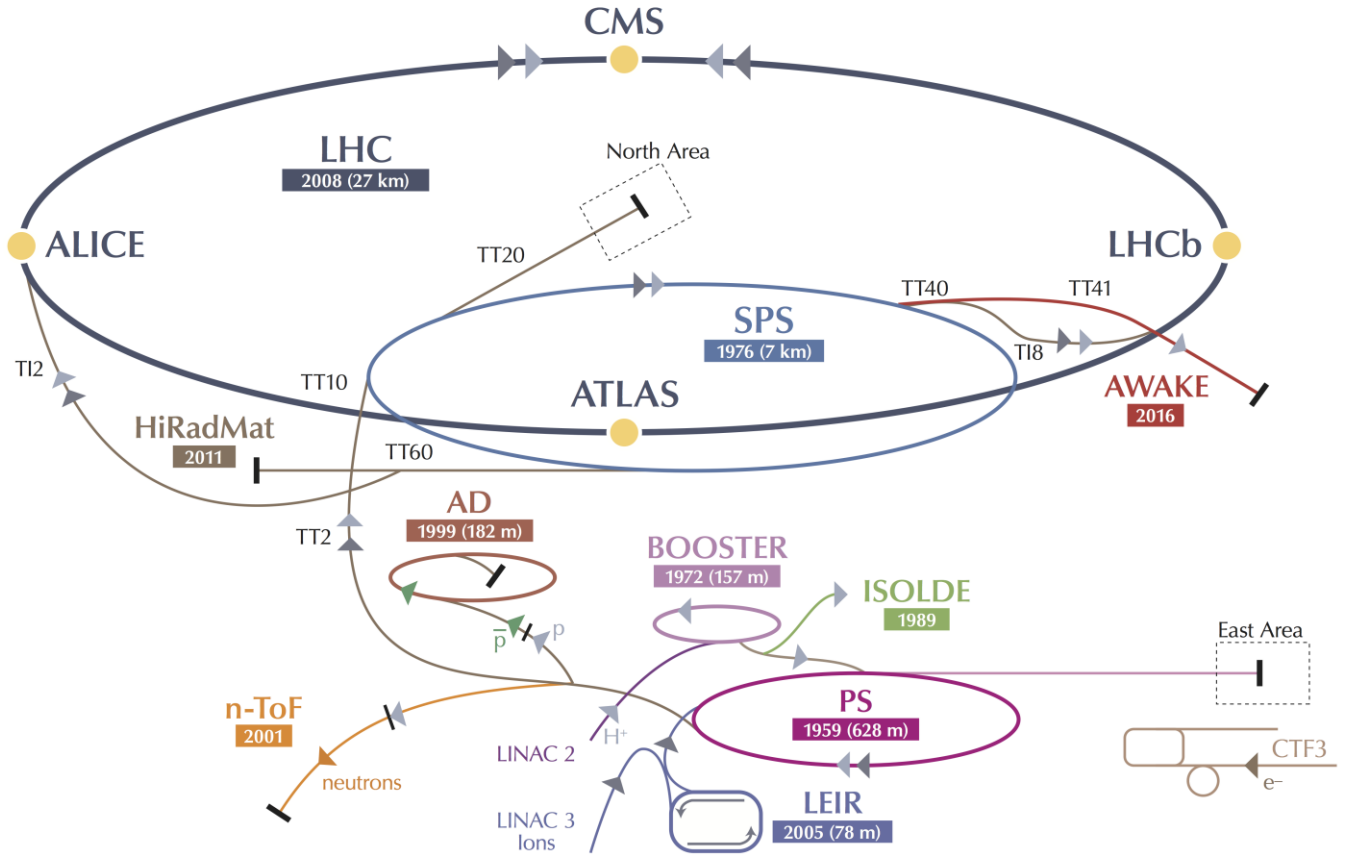
Observe and record the results of these collisions



The Large Hadron Collider (LHC)

- Largest machine in the world
  - 27km, 6000+ superconducting magnets
- 600 million collisions per second
  - Generating approximately one petabyte of data per second
- 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

The CERN Accelerator Complex



p (proton) ion neutrons p̄ (antiproton) electron proton/antiproton conversion

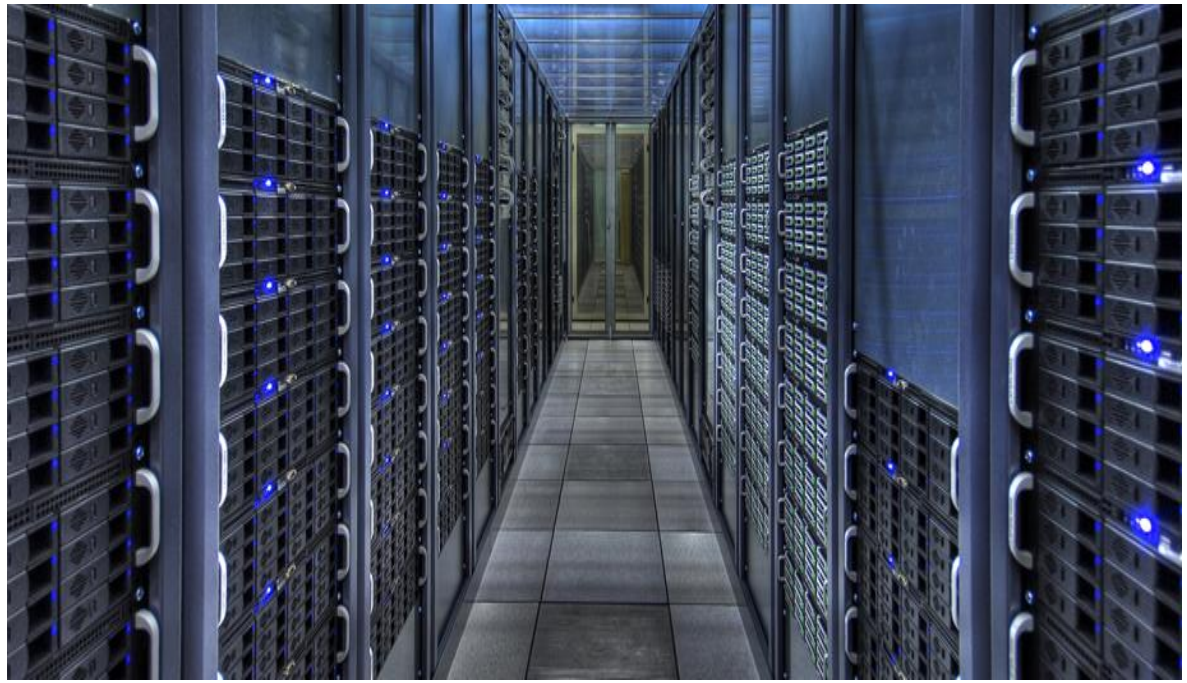
LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron

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

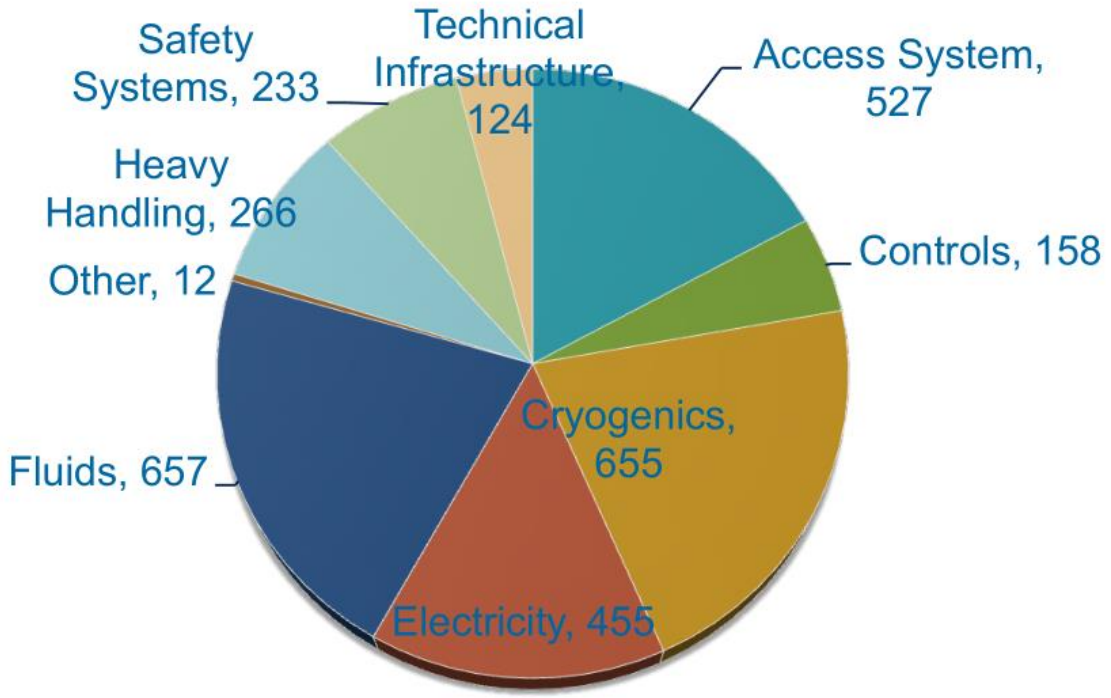
CERN is an extreme data environment

- Control and operations
  - Millions of sensors, signals
  - Large number of control devices
  - Equipment
- Monitoring and logging
- Supporting IT infrastructure
  - Databases
  - Network
  - Services
- CERN has great monitoring and logging systems
  - Large amount of data has been stored over years



Data Analytics Challenges

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

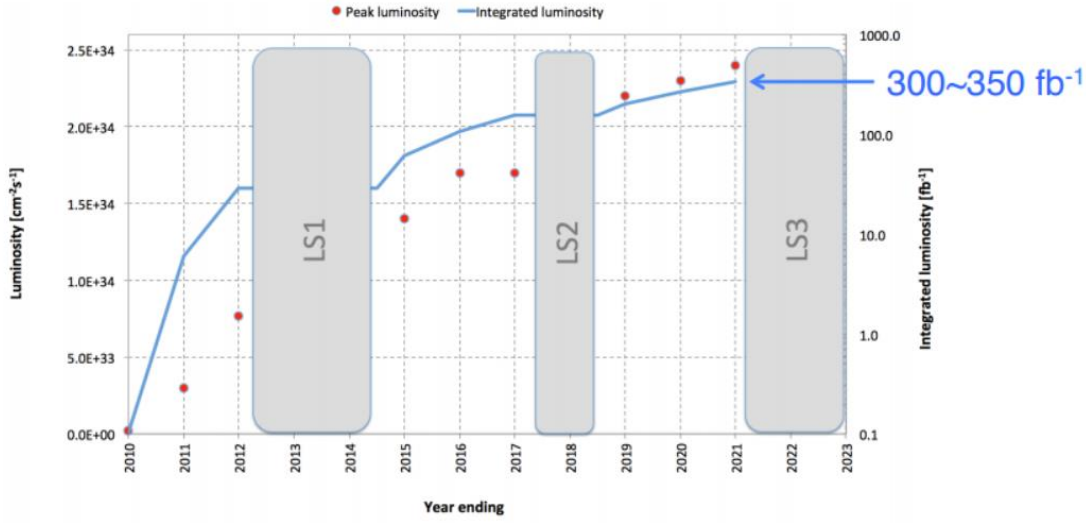


A look into the Future

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

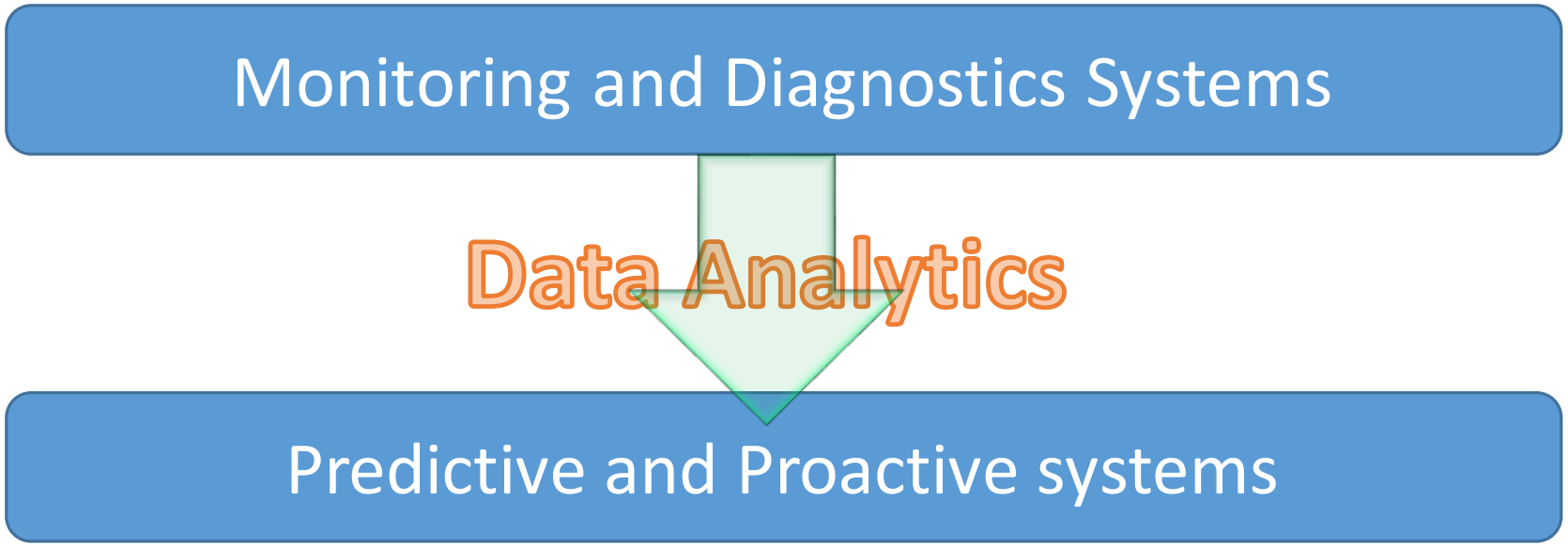
Table 17: HLT Rate Evolution

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)

The objective – Improve our systems



openlab 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

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

CERN openlab

Partners

HUAWEI

intel

ORACLE

SIEMENS

Contributors

rackspace
the open cloud company

Associates

Yandex

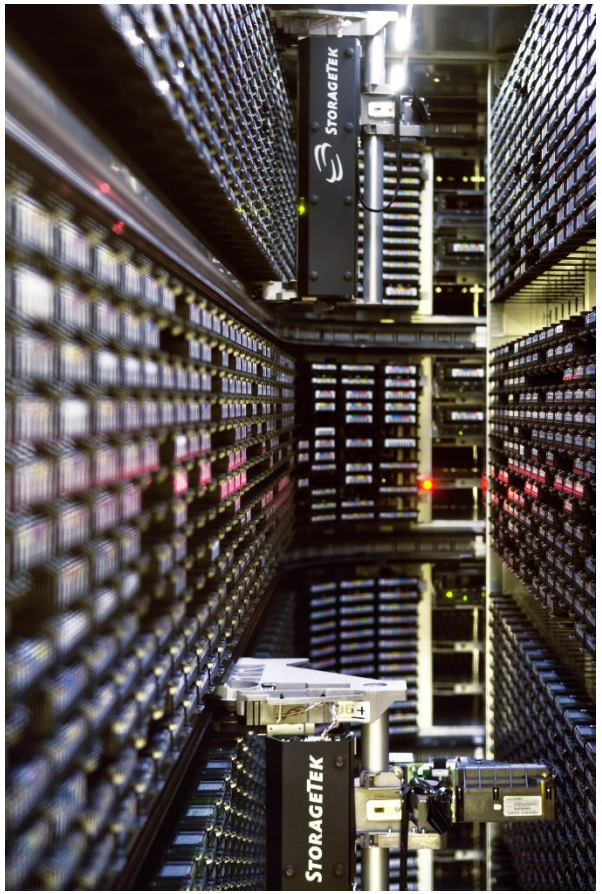
Data Analytics Use Cases

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

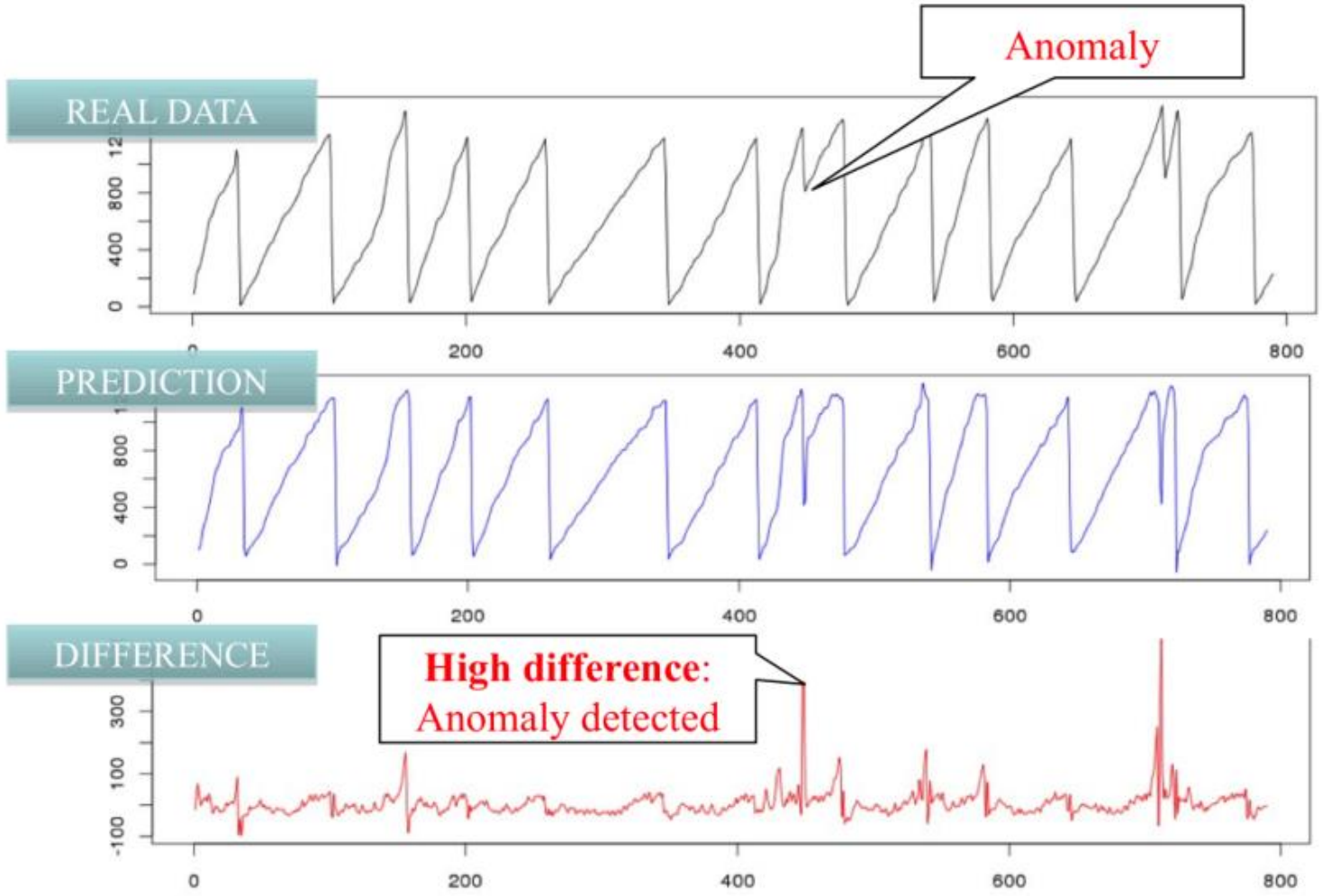
CASTOR

CERN Advanced STORage manager



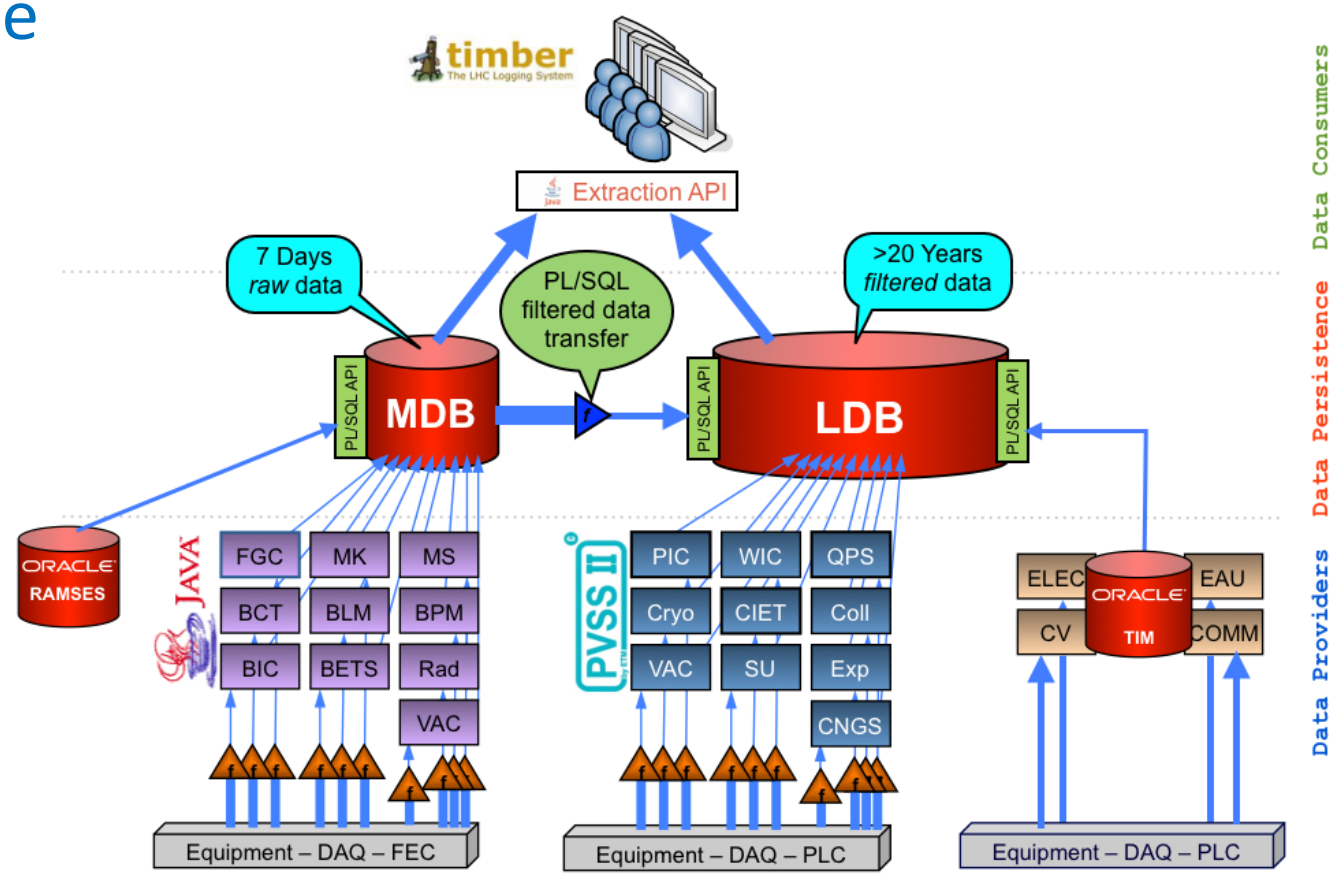
Use Case: CERN Advance Storage Manager (CASTOR)

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



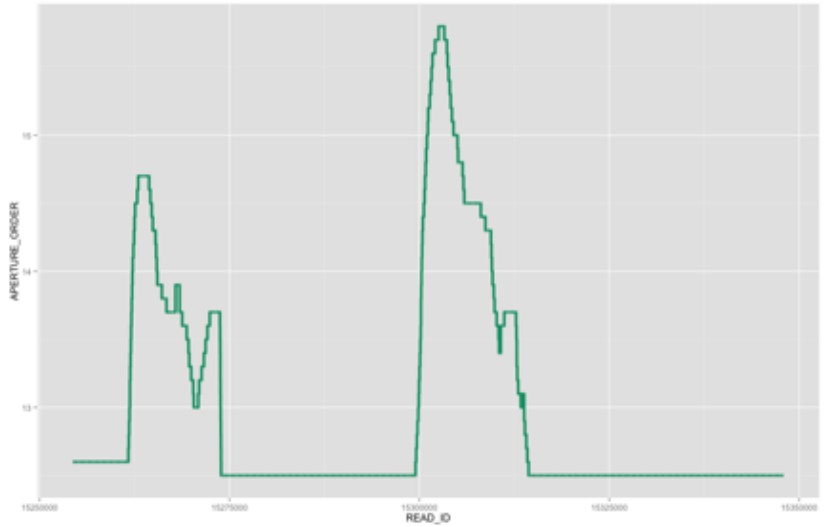
Use Case: Operation and Control Systems

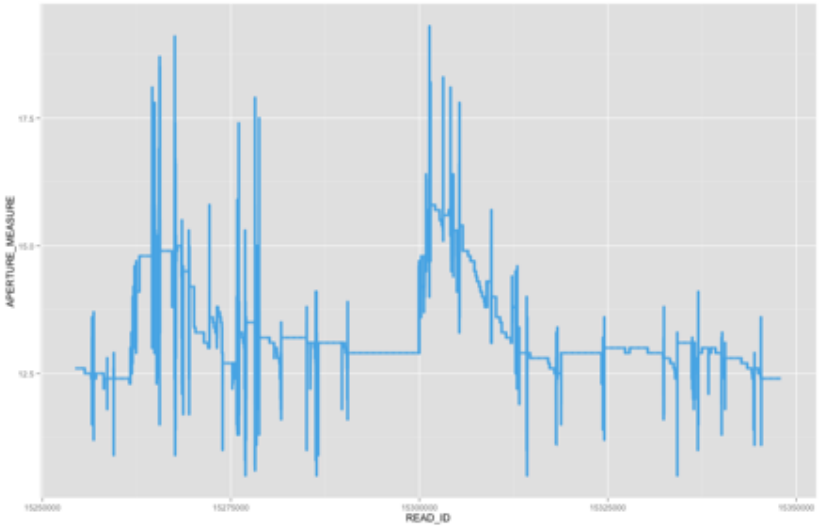
- LHC Accelerator Logging Service
  - Around 1 million signals
    - Temperatures, electrical currents
    - Magnetic field strengths
    - Vacuum pressures, etc.
- Control system Health
  - Gas Breakdown
  - Vacuum
  - Machine Protection
- Predictive maintenance

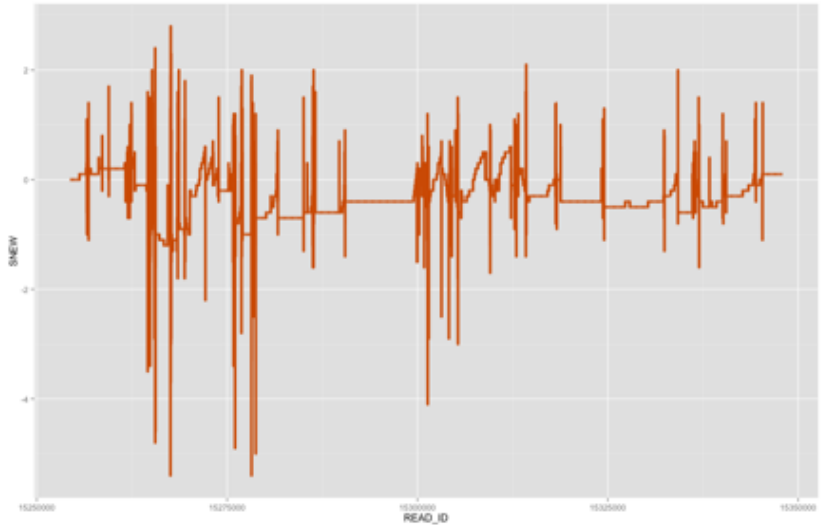


Use Case: Cryogenics Faulty Valves Detection

- 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

Data Discovery for Accelerator Complex Operations

- Interactive and visual analytics by end users
  - Flexible and easy to use but powerful
  - Analyze information of any type and any source
  - Get **new insights** from data
- Electronic Logbook
  - Log of events in the accelerator complex

Statistics for the eLogbook: PS From: 20120815 Period: Morning To: 20120817 Period: Morning

Start Statistics ...

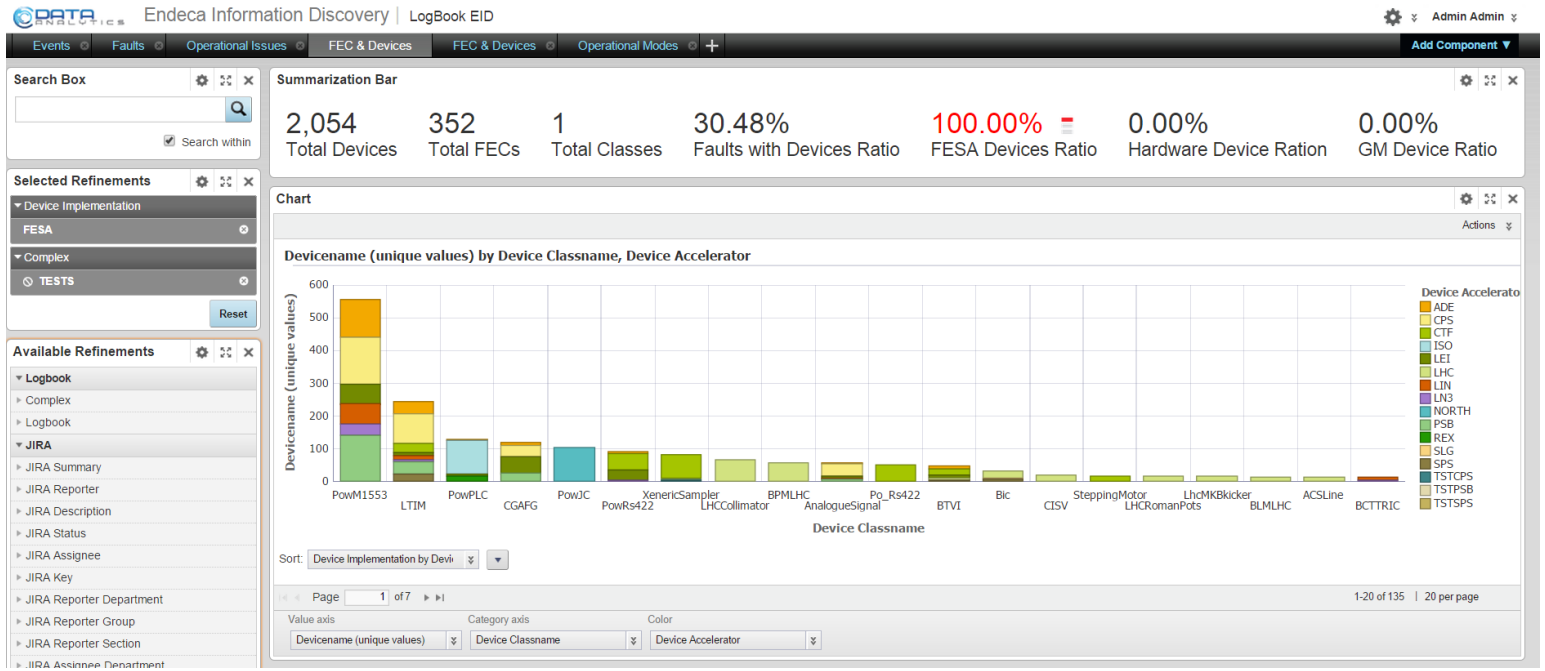
Availabilities			
Lines	In Super Cycle	In Fault	Availabilities (%)
AD	55 [h]	7 [h] 57 [min] 37[s]	85%
EASTA	55 [h]	7 [h] 57 [min] 37[s]	85%
EASTB	55 [h]	7 [h] 57 [min] 37[s]	85%
SFTPRO	6 [h] 54 [min] 29[s]	5 [h] 45 [min] 18[s]	16%
EASTC	55 [h]	7 [h] 57 [min] 37[s]	85%
CNGS	55 [h]	10 [h] 22 [min] 58[s]	81%
LHCPROBE	55 [h]	7 [h] 57 [min] 37[s]	85%
I_LHC	31 [h] 31 [min] 33[s]	4 [h] 37 [min] 40[s]	85%
LHC	55 [h]	7 [h] 57 [min] 37[s]	85%
TOTAL	423 [h] 26 [min] 02[s]	68 [h] 31 [min] 38[s]	83%

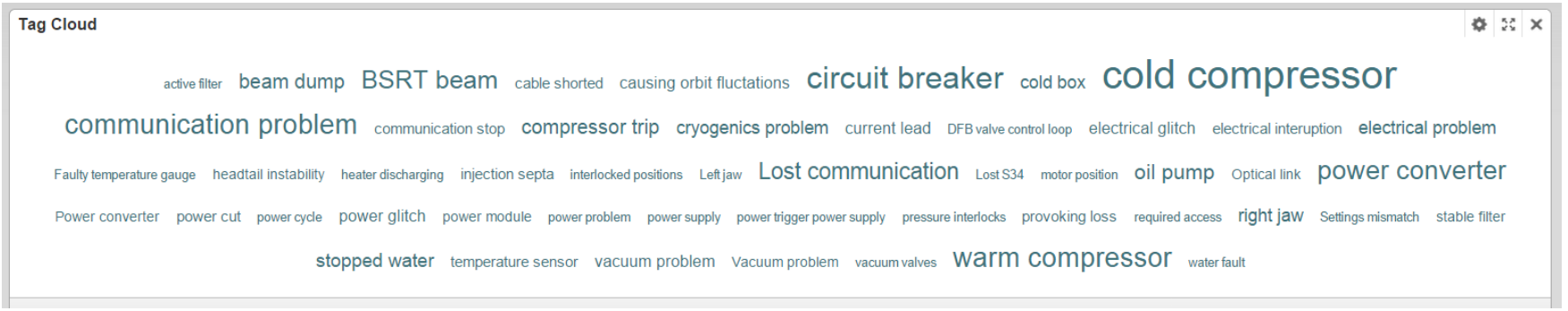
Systems for AD		
GROUP NAME	FAULT NAME	DURATION
PS	Power supply	3 [h] 47 [min] 34[s]
PS	RF	3 [h] 13 [min] 33[s]

4	23:42	SUP	Global Post Mortem Event Event Timestamp: 10/06/12 23:42:39.163 Fill Number: 2718 Accelerator / beam mode: PROTON PHYSICS / STABLE BEAMS Energy: 4000080 [MeV] Intensity B1/B2: 15509 / 14217 [e^10 charges] Event Category / Classification: PROGRAMMED_DUMP / MULTIPLE_SYSTEM_DUMP First BIC input Triggered: First USR_PERMIT change: Ch 1-Programable Dump b1: A T -> F on CIB.CCR.LHC.B1
5	23:42	SUP	Global Post Mortem Event Confirmation Dump Classification: Programmed Dump Operator / Comment: papotti / End of physics fill, clean dump.
6	23:42	SUP	**BEAM MODE > BEAM DUMP** LHC RUN CTRL: BEAM MODE changed to BEAM DUMP
7	23:42	SUP	**BEAM MODE > BEAM DUMP** LHC RUN CTRL: BEAM MODE changed to BEAM DUMP
8	23:42	SUP	ELOGBOOK: STARTING B1 MKISS
9	23:43	SUP	ELOGBOOK: STARTING B2 MKISS
10	23:44	SUP	LHC SEQ: beam dump handshake closed; LHC=STANBY, EXP=VETO
11	23:44	SUP	LHC SEQ: MCS checks finished
12	23:45	SUP	LHC SEQ: SMP pre-operational checks finished
13	23:45	SUP	LHC SEQ: BIS pre-operational checks finished
14	23:48	SUP	**BEAM MODE > RAMP DOWN** LHC RUN CTRL: BEAM MODE changed to RAMP DOWN
15	23:48	SUP	LHC SEQ: BPMLHC calibration finished. Overall result: SUCCESS Chosen bunch spacing: (B1 & B2) BUNCH_50NSEC (manually chosen) (For more details see BI-LHC ELogBook)

Data Discovery for Accelerator Complex Operations

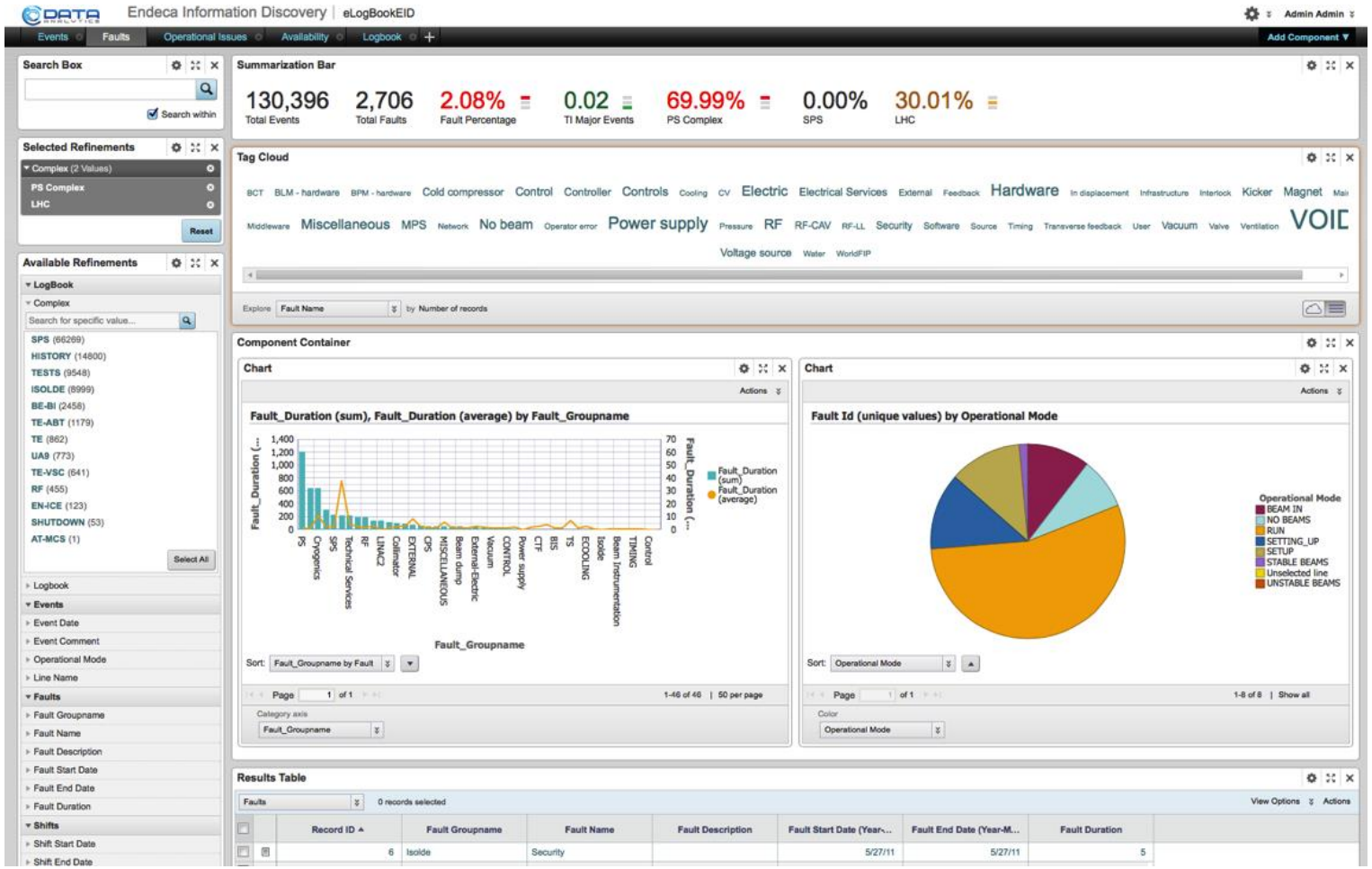
- Integrating Data Sources
  - Electronic Logbook
  - Controls Configuration DB
  - JIRA (JSON)
- Text analysis
  - Extract most relevant terms
  - Entity extraction





Data Discovery for Accelerator Complex Operations

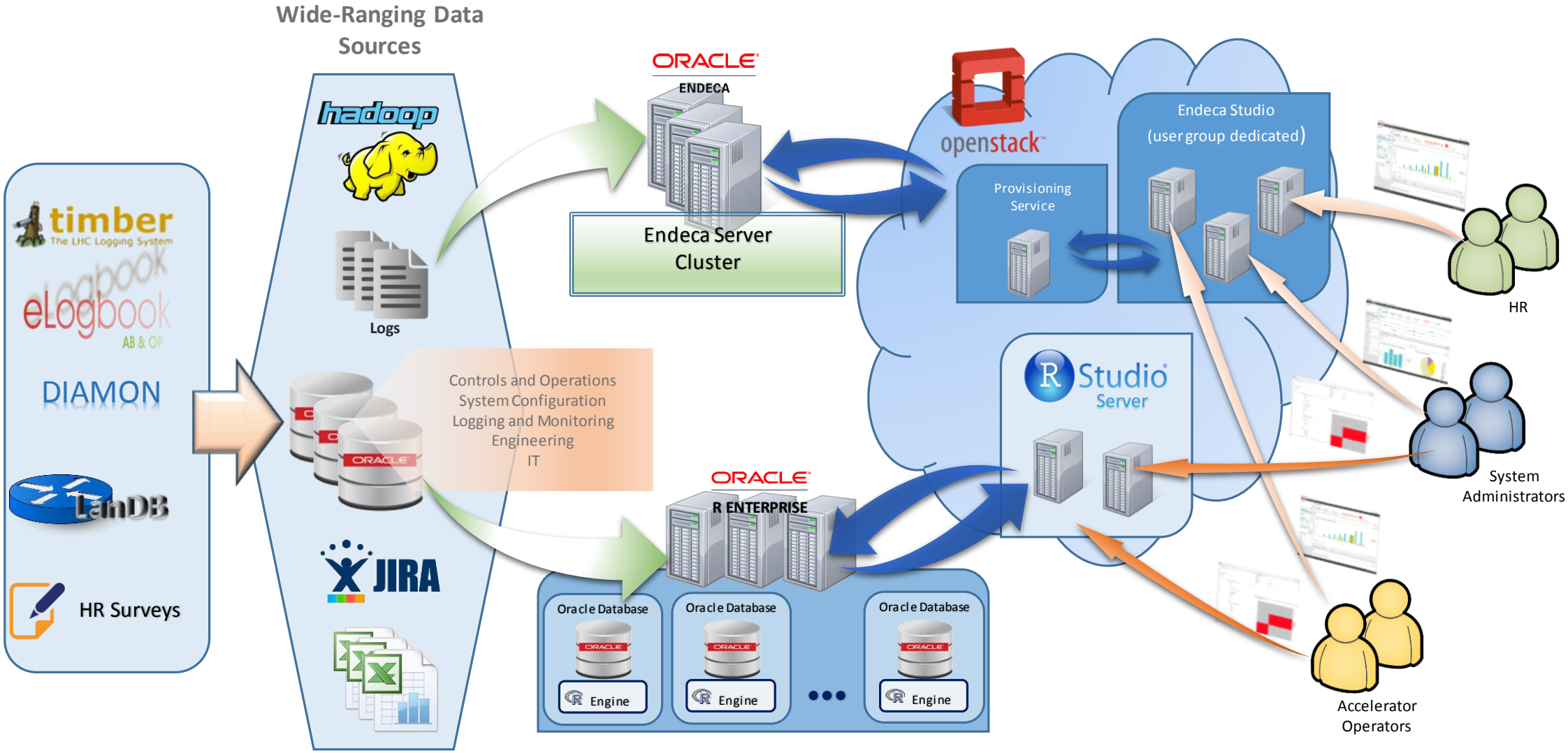
- Better LHC Operations
  - Events Analysis
  - Correlate Information
    - Fault Tracking System
    - Operational Modes
    - Operational Issues
    - Control Equipment



Data Discovery for Accelerator Complex Operations

- Many different Applications
  - Controls and Operations
    - Accelerator Fault Tracking
    - Diagnostics and Monitoring
  - IT Infrastructure Monitoring
    - Server logs analysis
    - Database latency
    - Accelerator Logbook
  - Human Resources

Infrastructure



www.cern.ch