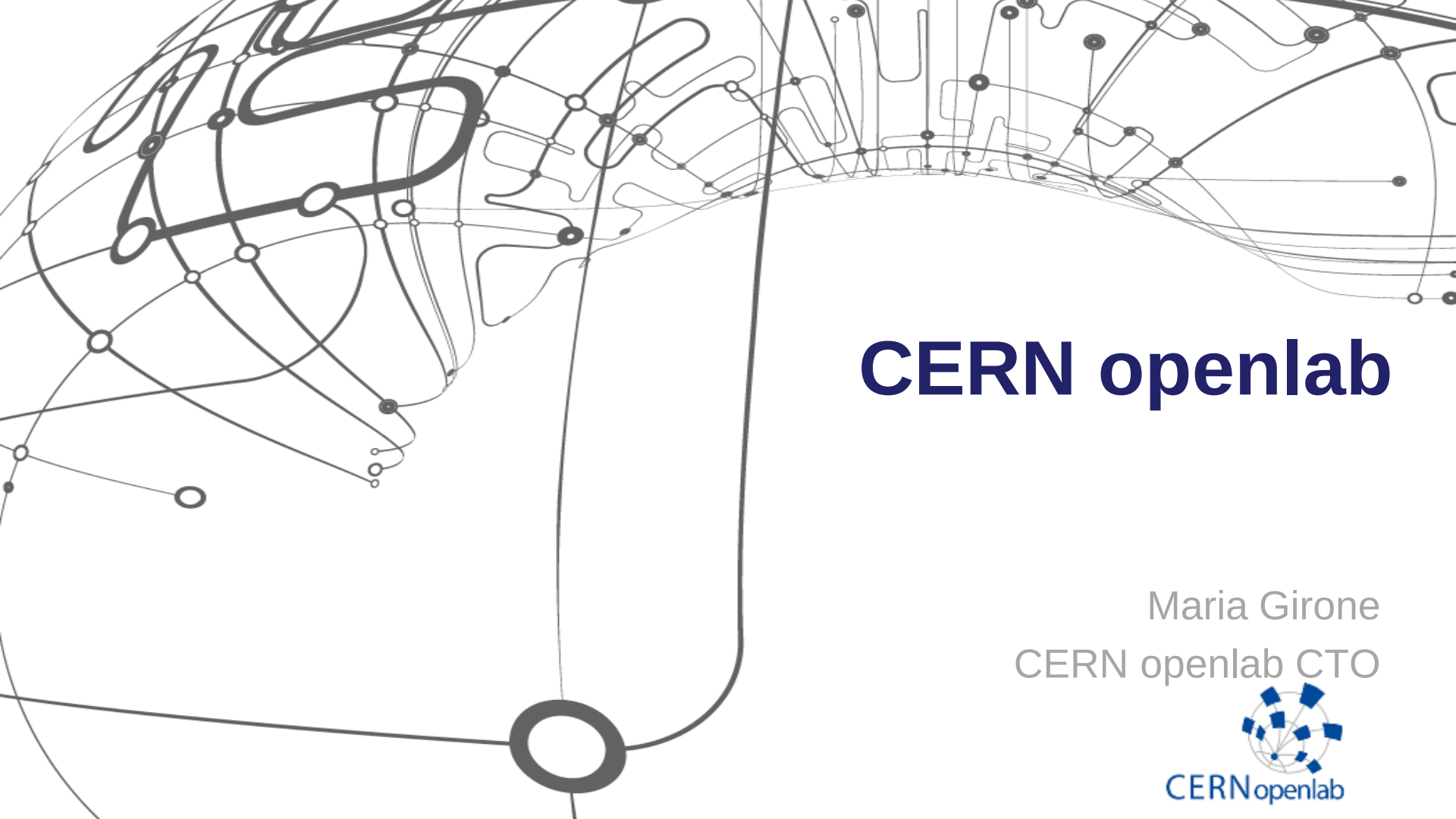


CERN openlab

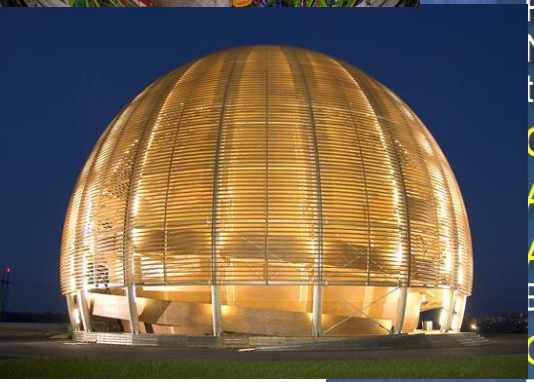
Maria Girone
CERN openlab CTO

CERN openlab

European Organization for Nuclear Research, founded in 1954 by 12 European countries

"Science for Peace"

~ 2300 staff
~ 1600 other paid personnel
10500 scientific users
udget (2014) ~1000 MCHF



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

Candidate for Accession: Romania

Associate Member in Pre-Stage to Membership: Serbia

Applicant States for Membership or Associate Membership: Brazil, Croatia, Cyprus, Pakistan, Russia, Slovenia, Turkey, Ukraine

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

➔ Interfacing between fundamental science and key technological developments

➔ CERN Technologies and applications

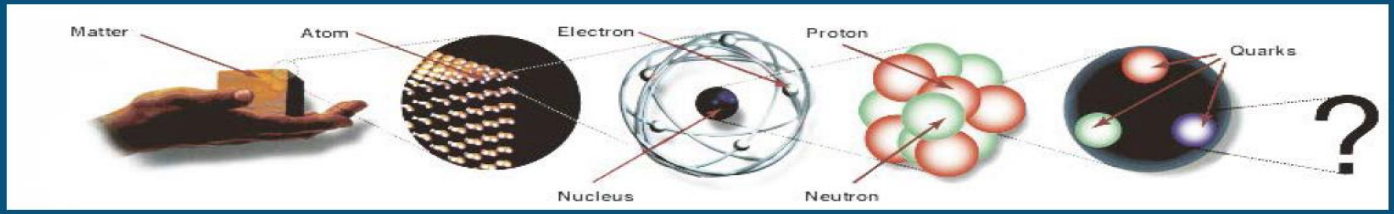
Accelerating particle beams

Detecting particles

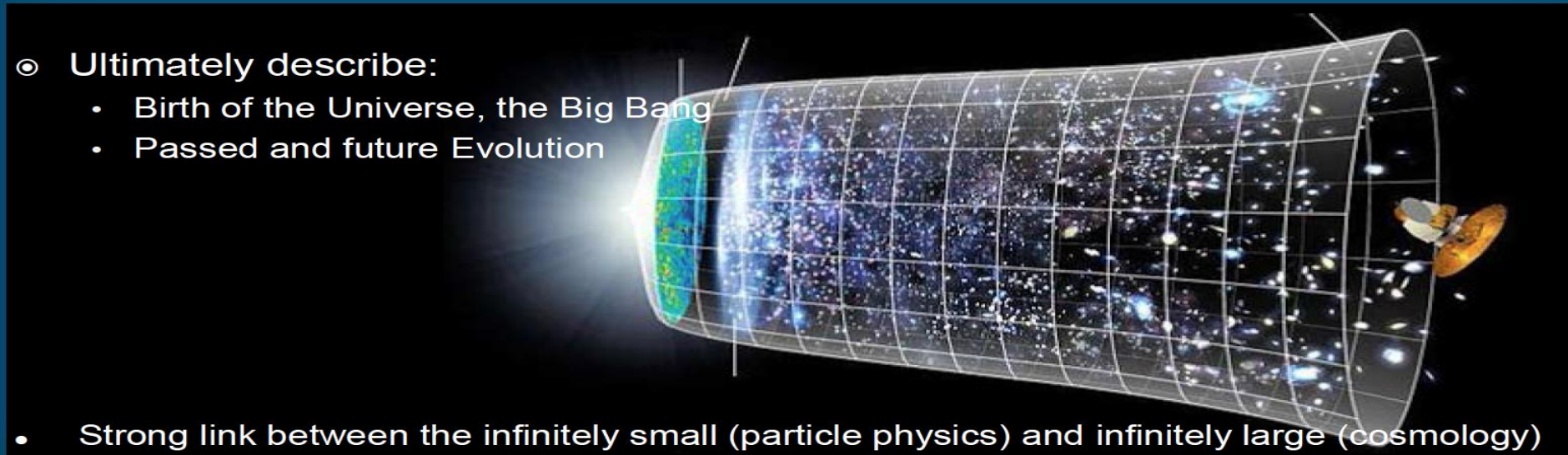
Large-scale computing (Grid)

Number 1: “Particle physics”

- Quest to understand:
  - Fundamental constituents of matter - Matter particles
  - Interactions with which particles act on each other - Interactions
  - Particles propagating the interactions - Messenger particles



- Ultimately describe:
  - Birth of the Universe, the Big Bang
  - Passed and future Evolution



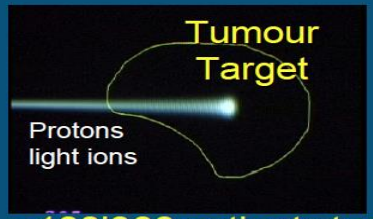
- Strong link between the infinitely small (particle physics) and infinitely large (cosmology)

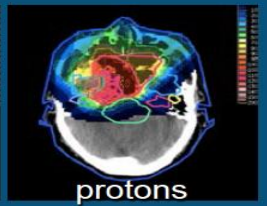
Number 2: Innovation

Medical applications

Accelerating particle beams
~30'000 accelerators worldwide
~17'000 used for medicine

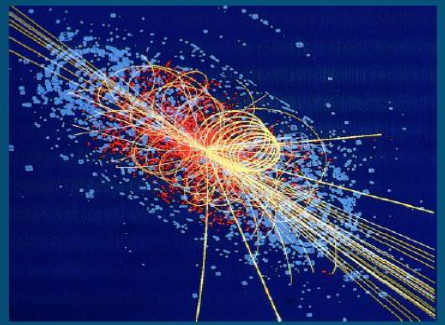
Hadron Therapy





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)

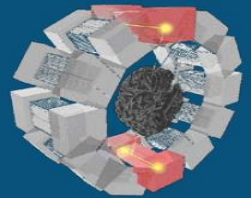


Detecting particles

Imaging

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

PET Scanner

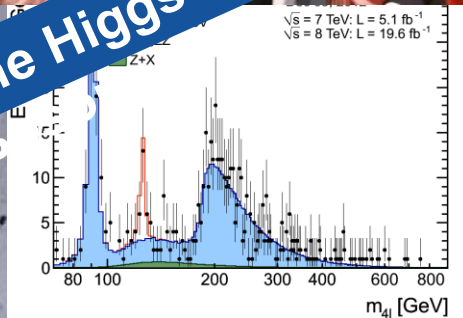


CERN: A UNIQUE ENVIRONMENT TO PUSH TECHNOLOGIES TO THEIR LIMITS

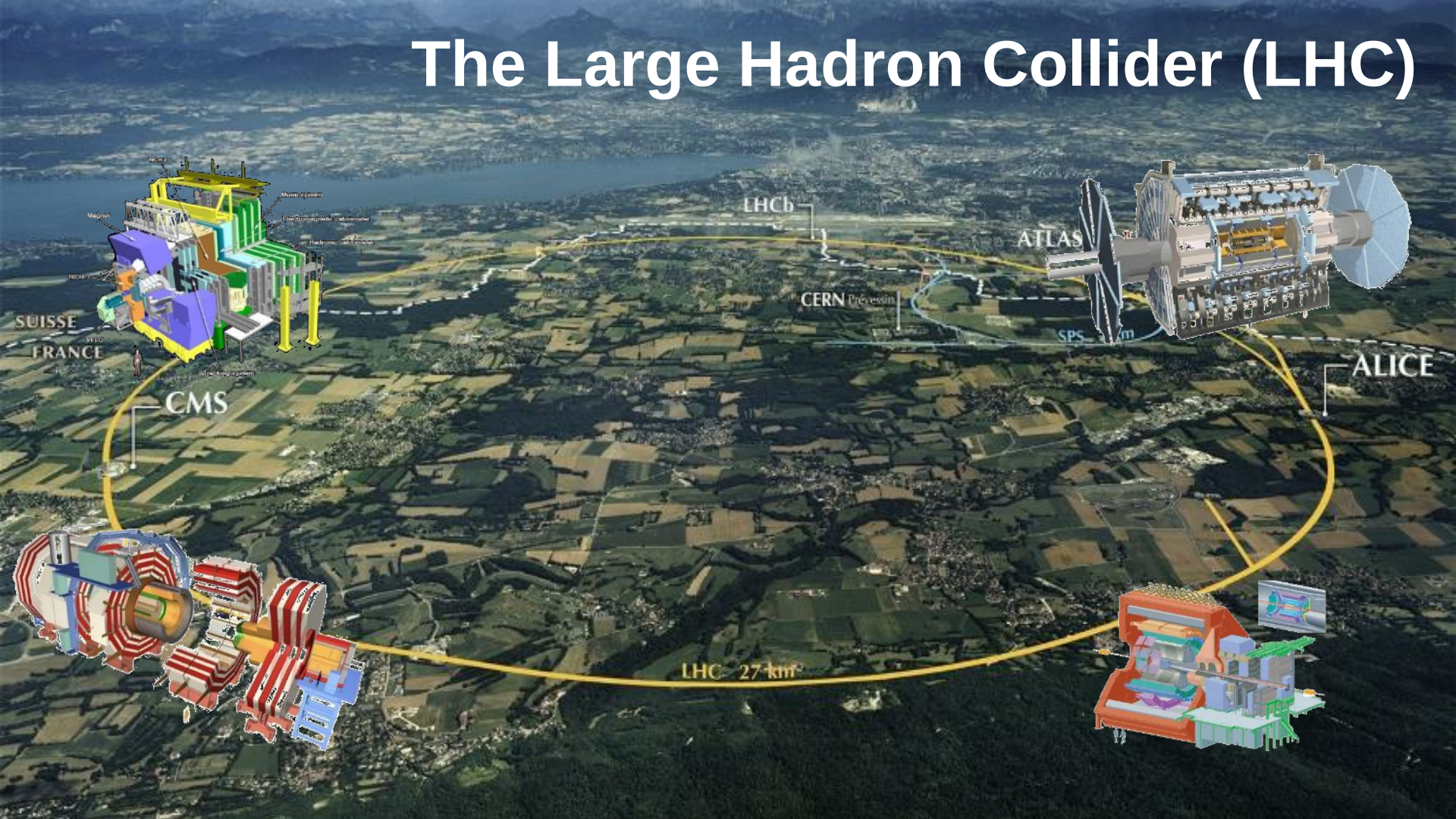
In its 60 year life CERN has made some of the important discoveries in particle physics

- Observation of the W and Z Bosons
- The number of neutrino families
- The Higgs Boson Discovery

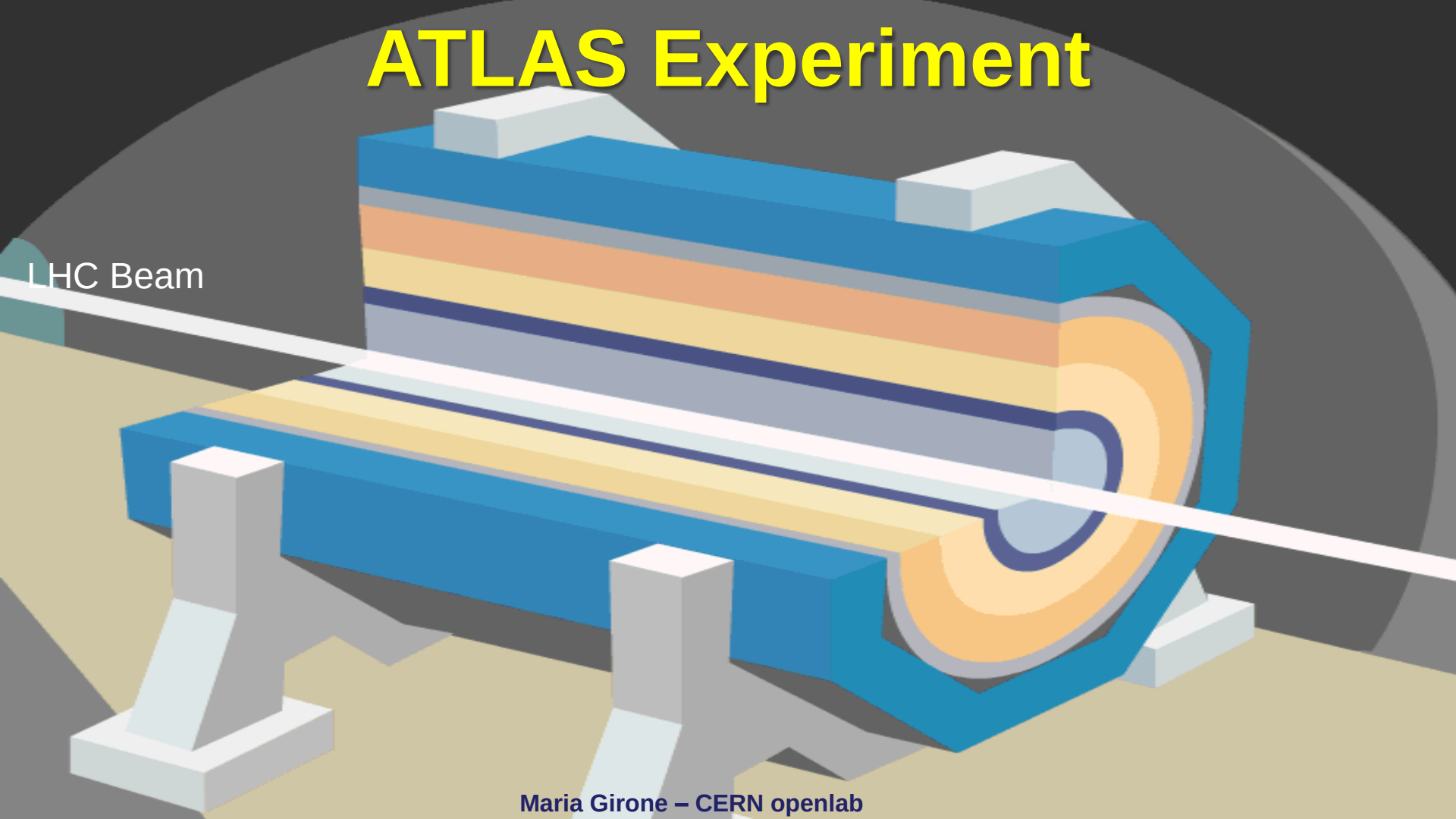
The Higgs B



The Large Hadron Collider (LHC)



ATLAS Experiment



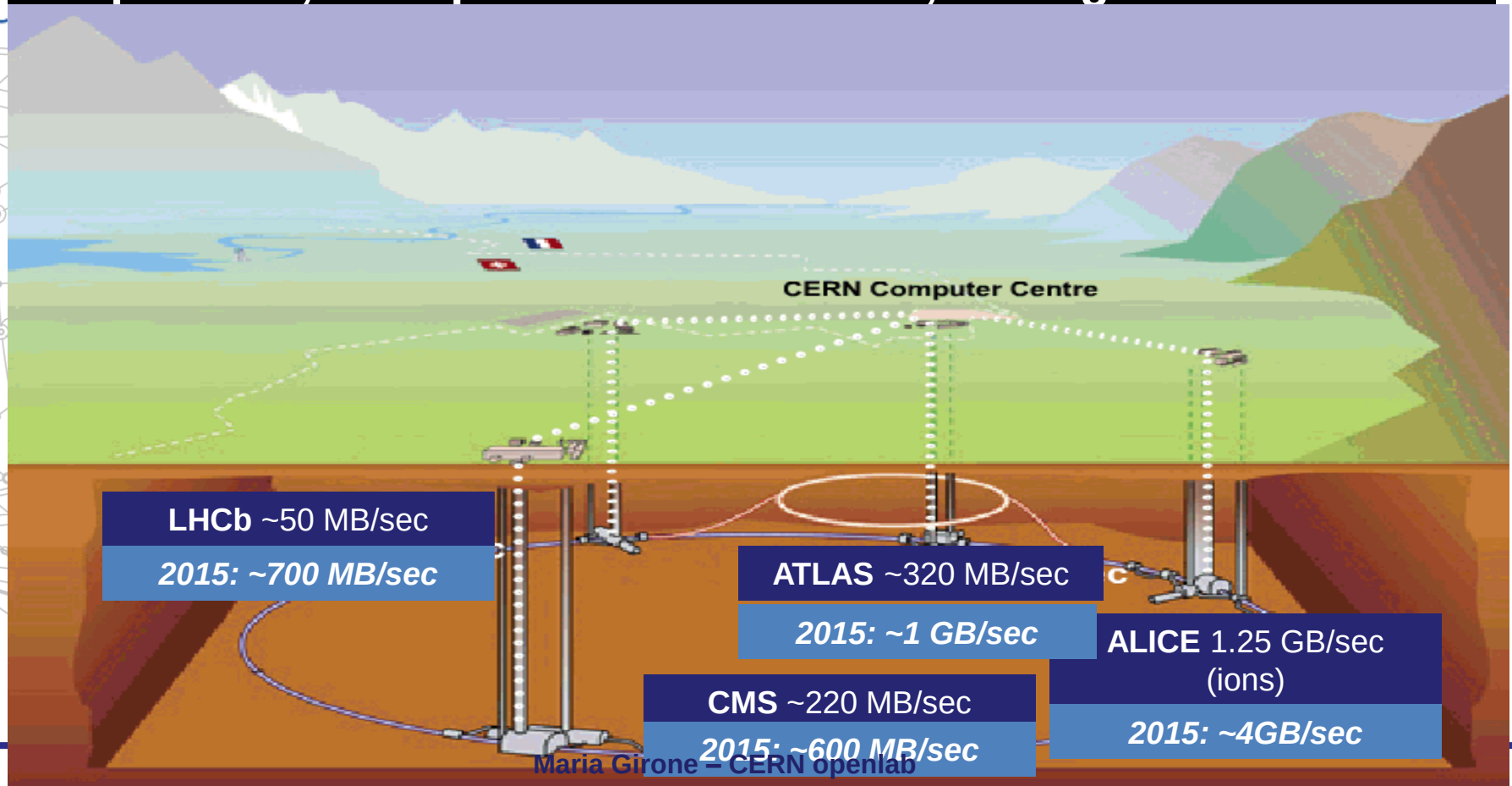
Data from ATLAS

Reduction factor of 1 million.

1 PB/sec from all sub-detectors

1 GB/sec raw data sent to Data Centre

CERN Computer Centre (Tier-0): Acquisition, First pass reconstruction, Storage & Distribution



1 PB/s of data generated by the detectors

Up to 30 PB/year of stored data

A distributed computing infrastructure of half a million cores working 24/7

An average of 40M jobs/month

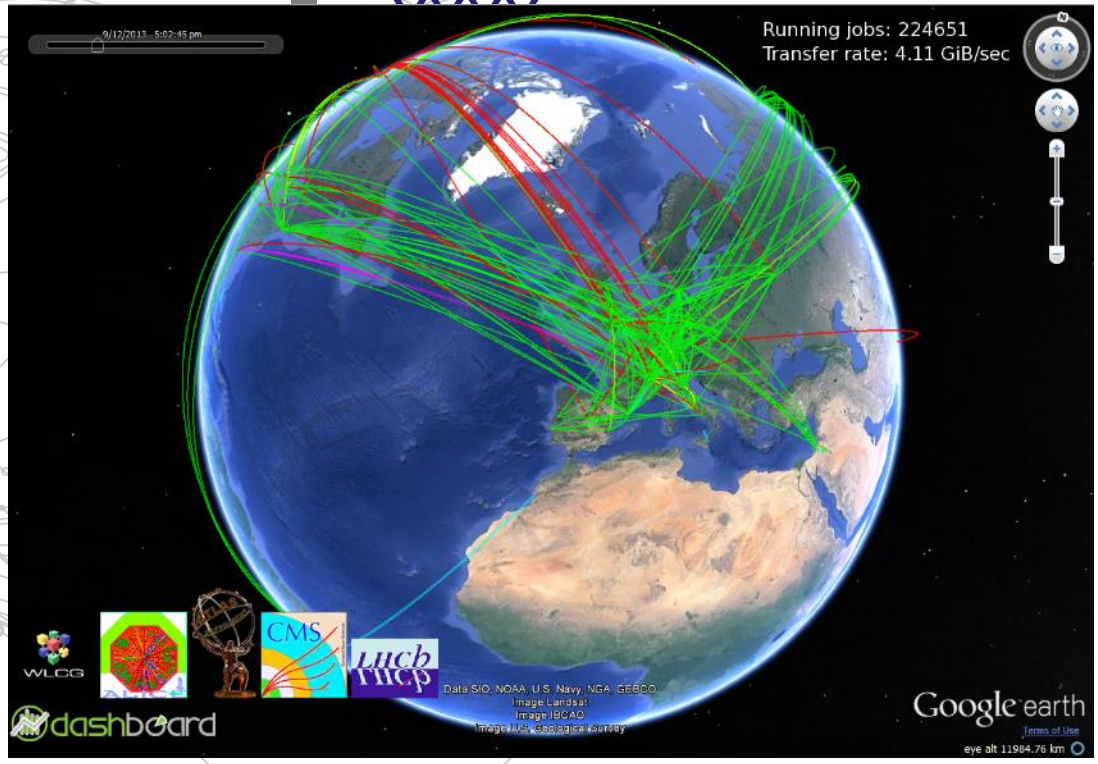
An continuous data transfer rate of 6 GB/s (600TB/day) across the Worldwide LHC Grid (WLCG)

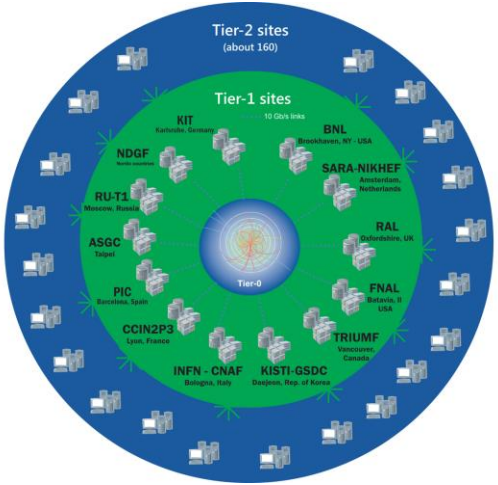
A sample equivalent to the accumulated data/simulation of the 10 year LEP program is produced 5 times a day

Would put us amongst the top Supercomputers if centrally placed

More than 100PB moved and accessed by 10k people

Worldwide LHC Computing Grid





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

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

Tier-2 (68 Federations, ~140 centres):
- Simulation
- End-user analysis

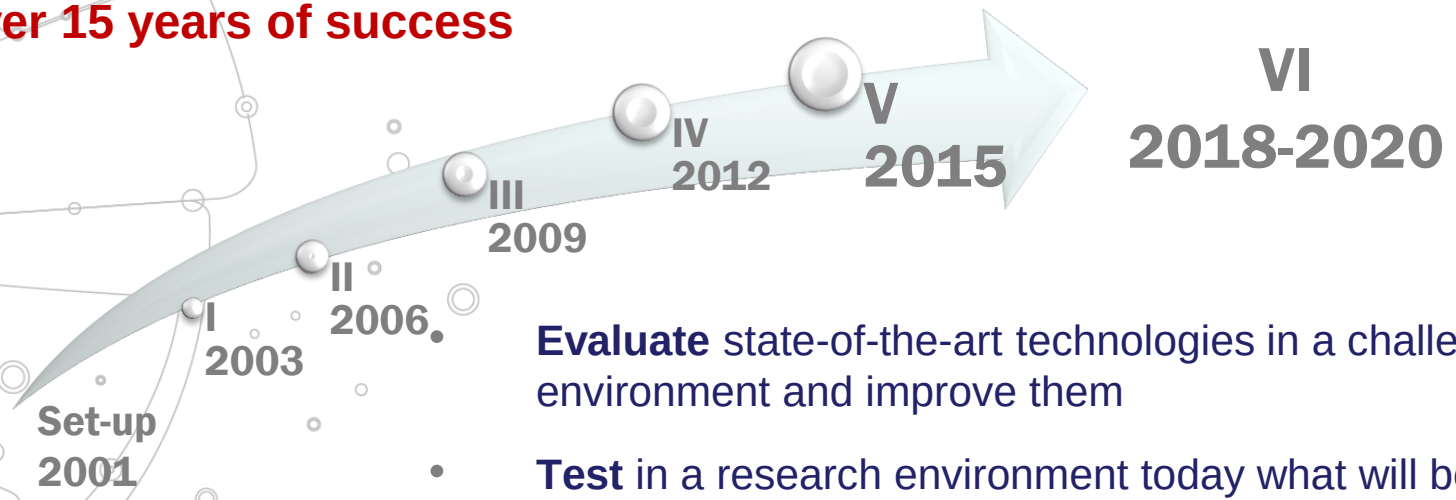
- 525,000 cores
- 450 PB

CERN openlab

- CERN openlab has been created to support the computing and data management goals set by the LHC programme
  - 15 years of innovative projects between CERN groups and experiments and leading IT companies
- In its phase V, CERN openlab is working to solve some of the key technical challenges facing the LHC in Run3 and Run4
  - Mutual benefit for industry and research communities
  - Collaborating with other research communities beyond HEP
- Ever-increasing interest in CERN openlab
  - well established mechanism of partnership between industry and research communities
  - a path to common developments for future challenges

15 Years of Successful Collaboration

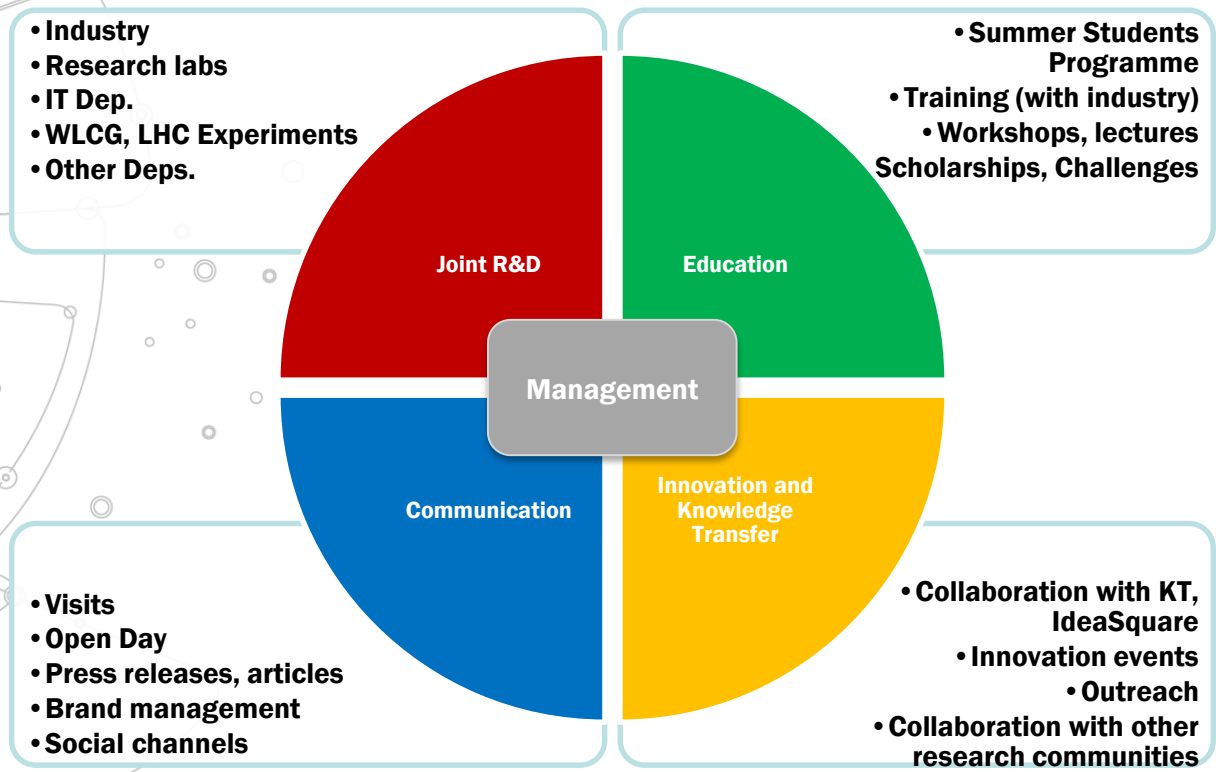
A science – industry partnership to drive R&D and innovation with over 15 years of success



- **Evaluate** state-of-the-art technologies in a challenging environment and improve them
- **Test** in a research environment today what will be used in many business sectors tomorrow
- **Train** next generation of engineers/employees
- **Disseminate** results and outreach to new audiences

CERN openlab in a nutshell

A science – industry partnership to drive R&D and innovation with over 15 years of success

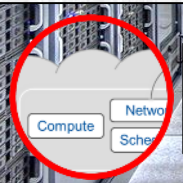


CERN openlab V Research Areas

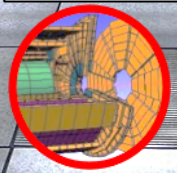
Data acquisition and filtering
Collecting data

Networks and connectivity
Connecting resources

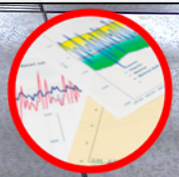
Data storage architectures
Storing and serving data



Compute management and provisioning (cloud)
Managing resources for processing



Computing platforms, data analysis, simulation
Improving processing and code efficiency



Data analytics/Machine Learning
Extracting information

Medical applications

Phase V Activities and Collaborations

Research Areas

- Data acquisition
- Computing platforms
- Data storage architectures
- Compute provisioning and management
- Networks
- Data analytics

In 2016, 27 FTEs, 17 ongoing projects

~40 openlab summer students

2016 Collaboration members

Ongoing openlab V Projects

- Data acquisition
- Networks and connectivity
- Data storage architectures
- Compute provisioning and management
- Computing platforms
- Data analytics

intel High-Throughput Computing Collaboration (HTCC)
IDT RapidIO for data acquisition

BROCADE Flow Optimizer Software

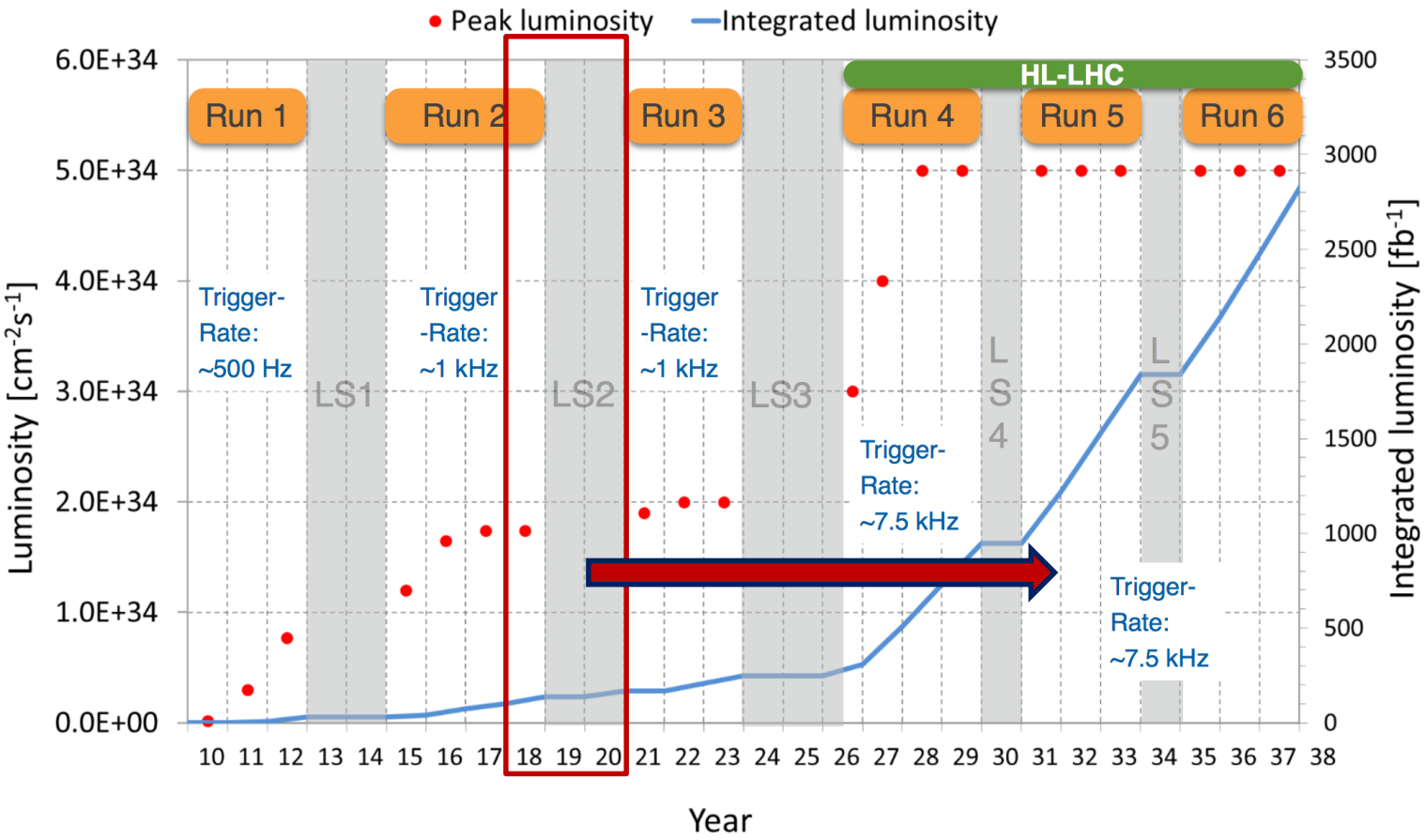
ORACLE Database Technology and Monitoring
SEAGATE Alternative Storage Architecture Kinetic
COMTRADE EOS Productisation

rackspace Containers
ORACLE Database Cloud
ORACLE Java EE

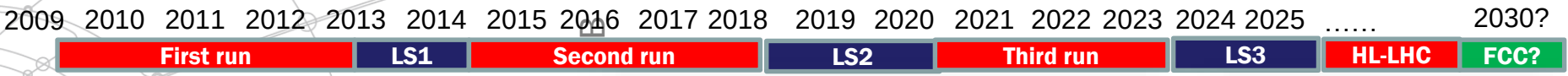
intel Code Modernization
HUAWEI ARM porting, optimization and benchmarking

intel Data Analytics
ORACLE Analytics-as-a-service
SIEMENS Industrial Control and Monitoring
Yandex Data popularity and anomaly detection

LHC Schedule

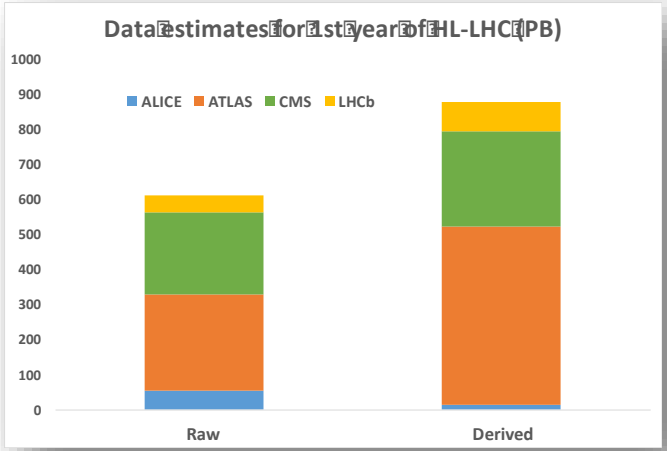


LHC Run3 and Run4 Scale and Challenges

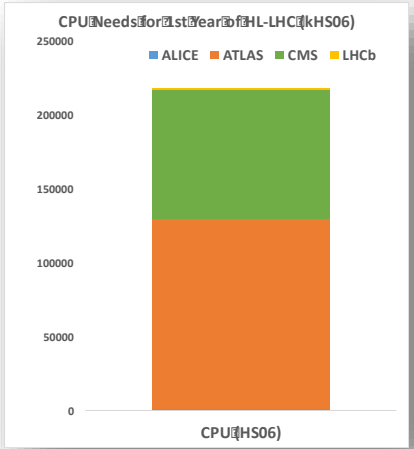


- Raw data volume for LHC increases exponentially and with it processing and analysis load
- Technology at ~20%/year will bring x6-10 in 10-11 years
- Estimates of resource needs at HL-LHC x10 above what is realistic to expect from technology with reasonably constant cost

Technology revolutions are needed



Data:
- Raw 2016: 50 PB → 2027: 600 PB
- Derived (1 copy): 2016: 80 PB → 2027: 900 PB



CPU:
- x60 from 2016

Courtesy of I. Bird

Looking for breakthroughs

- The LHC Experiments are looking in a variety of areas to help close the gap in resources (perform the same tasks with less total computing)
- **Data Center Technologies**
- **Code Modernization**
- **Data Analytics**
  - Improving operations and maintenance of the accelerator, detectors, and the computing infrastructure
  - Improve how we analyze the events collected, to perform the same level of analysis on a much larger collection of events
- **Machine Learning**
  - Improve efficiency for using resources, reduce CPU required for reconstruction and identification, improve the event handling

Data Analytics and Machine Learning

- The community is opening to modern techniques in the field and turns to industry who is leading the way
  - Interest from new researchers to learn new skills
- CERN openlab LHC experiment plans and needs in the areas of machine learning and data analytics presented in April 2016 at the CERN openlab DA and ML workshop https://indico.cern.ch/event/514434/timetable/#20160429.detailed
  - Large presence of leading industries at the workshop
- Another workshop will be held on April 27th 2017
- Proposed Proof of Concepts (POCs) are around
  - Streamlining analysis access
  - Event Categorization and Triggering
  - Anomaly Detection and Predictive Maintenance
  - Resource Infrastructure Optimization
  - Physics Object Identification (particle reconstruction)

The Intel Big Data Analytics Project

- The project has four use-cases
- **CERN/CMS (with Fermilab) : CMS Physic Data reduction Center**
  - Read through petabytes of data and produce a sample of reduced data based on potentially complicated user queries on the time scale of hours and not weeks as it currently requires
- **CERN/LHCCO – Control Data for LHC**
  - Collect control data from LHC SCADA systems in production and use them to create predictive analytics solution to anticipate possible LHC problem. Can be generalized

The Intel Big Data Analytics Project

- **CERN Root benchmarking**
  - Benchmark the Cloudera Kudu vertical data storage engine against the ROOT TTree object storage system in the areas of performance and features (e.g. schema evolution, compression, data format, etc.) using realistic LHC physics analysis data. Both systems are designed to store tables of structured data. Also aspects of data processing using **Spark and Root** will be compared and benchmarked for in memory and large out of memory data set queries.

- **CERN Medical Services – Epidemiologic Study (with SciPulse Foundation)**
  - Exploit the big-data developments pushed by the physics experiment, to explore whether and how the diverse data ecosystem already in use to monitor CERN's staff and users could inform new experimental hypothesis and, ultimately, speed up the evolution of safety and prevention practices at CERN.

Three Broad Areas of R&D

- **Data Center Technologies and Infrastructures**
  - Networks
  - Cloud Computing
  - Storage and databases
  - Data Center Architectures (disaggregation)
- **Computing Platforms and Software**
  - Architectures
  - Software modernization/acceleration
- **Data Analytics and Machine Learning**
  - Physics
  - Engineering (Control systems, infrastructure optimization)
  - Great interest from other communities

Three Topical Workshops

- Data Center Technologies and Infrastructures: March 1st
- Computing Platforms and Software: March 23rd
- Data Analytics and Machine Learning: April 27th

- **Whitepaper on Computing Challenges,** September 2017

- Q4 2017: renewal of FA and finalization of projects
- Jan 2018: start of Phase VI

Summer Student Programme

In 2016

- 1480+ applicants
- 39 selected students
- 16 lectures
- Visits to external labs and companies
- Lightning talks session
- 39 Technical reports

Outlook

- CERN openlab is a unique resource in HEP
  - A broad group of industry partners
  - A large R&D effort
  - An established framework for collaboration and innovation
- There is strong interest in industry to help develop techniques that will be critical to the success of the HL-LHC
  - Advanced data analytics, data movement and management, machine learning, software improvement, data center technologies, …
- The definition of phase VI is a great opportunity to align the activities of CERN openlab with the needs of the HEP community

EXECUTIVE CONTACT

Alberto Di Meglio, CERN openlab Head

alberto.di.meglio@cern.ch

TECHNICAL CONTACTS

Maria Girone, CERN openlab CTO

maria.girone@cern.ch

Fons Rademakers, CERN openlab CRO

fons.rademakers@cern.ch

COMMUNICATION CONTACTS

Andrew Purcell, CERN openlab Communication Officer

andrew.purcell@cern.ch

Mélissa Gaillard, IT Dep. Communication Officer

melissa.gaillard@cern.ch

ADMIN/FINANCE CONTACT

Kristina Gunne, CERN openlab Administration Officer

kristina.gunne@cern.ch

CERN openlab Framework

- **No Planned Changes in the Agreement**
  - Possibly just an extension of existing agreement for three more years
- **The Framework Agreement**
  - Signed between CERN and a company or research lab, valid until the end of the current phase, renewable
  - 25k CHF/year Membership Contribution
- **The Project Agreement**
  - One per each project, signed by all members taking part, can span across phases, contains the technical description of the work to be done
  - Costs and contributions to be negotiated based on duration and complexity of the project
    - Associate: only in-kind contributions (hardware, software licenses, expertise, etc.)
    - Contributor: same as Associate plus funds to recruit one dedicated Fellow at CERN (150k CHF/year)
    - Partner: same as Contributor plus funds to recruit additional effort as necessary (150k CHF/year per person)
- The membership contribution and each 150k CHF contribution for dedicated people include the cost to support summer students

CERN openlab Open Day

- High-visibility public event to promote the activities and opportunities created by CERN openlab and its projects
- **2015 Edition**: a kick-off event, traditional presentation-style format
- **2016 Edition**: exhibition-style showcase event with a few selected keynotes and interactive, open booths and stands, demos, short focused technical talks, etc.
- **2017 Edition**: again exhibition style, but focus on engaging more external communities, launch of WP on CC4SR