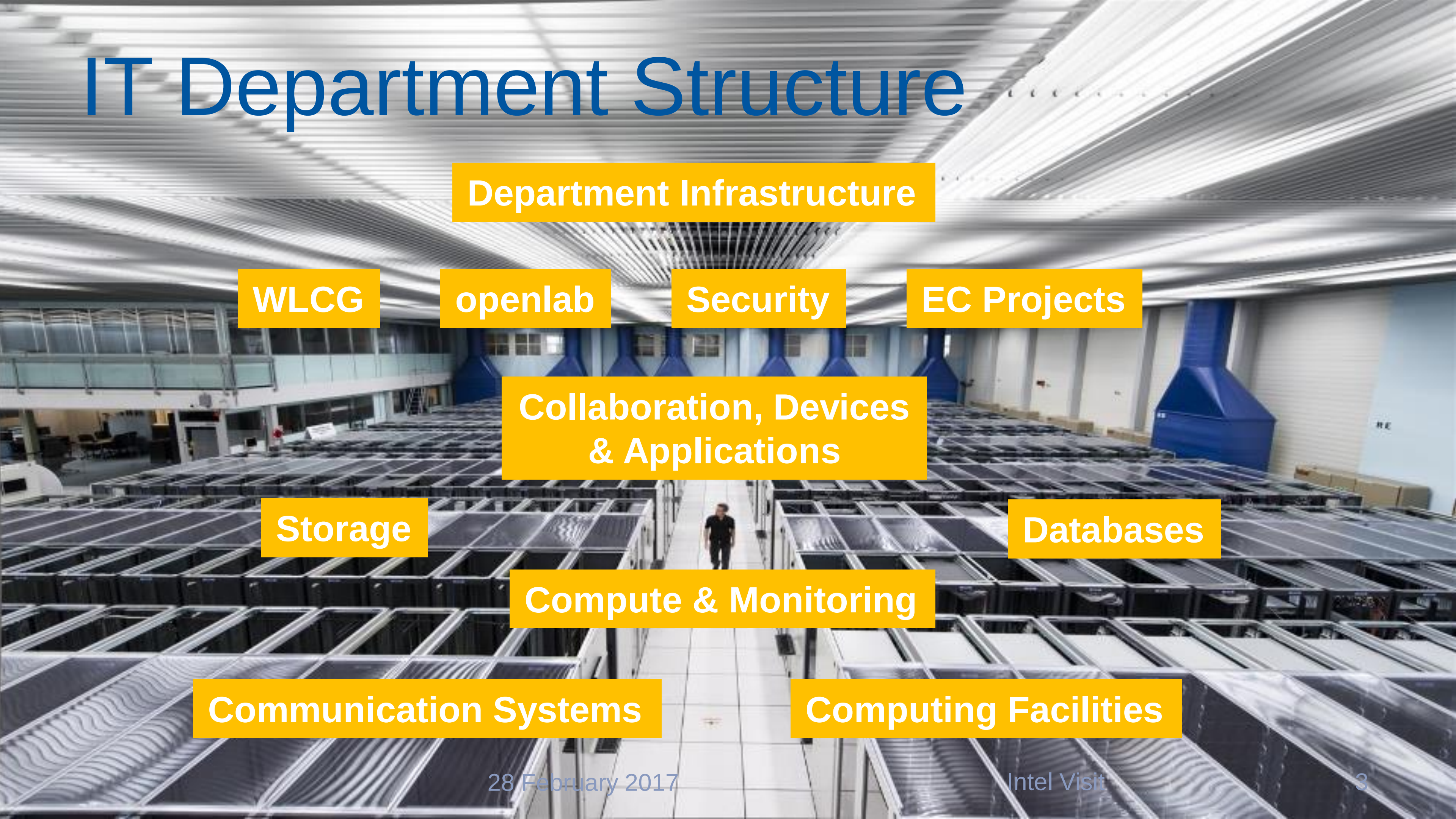
CERN

Vision for CERN IT Department

Frédéric Hemmer – IT Department Head



IT Department Structure

Department Infrastructure

WLCG

openlab

Security

EC Projects

Collaboration, Devices & Applications

Storage

Databases

Compute & Monitoring

Communication Systems

Computing Facilities

IT Services


IT Service Metrics 2016 Summary
Zoom Out
Jan 5, 2016 16:33:09 to Jan 2, 2017 23:59:59
Data Centre - Cores
Meyrin Wigner
Data Centre - Disks
Meyrin Wigner
Cloud - VMs
Virtual Machines
Network - Links to Wigner (Average IN Rate)
IN (GEANT) IN (T-Systems)
LXPLUS - Sessions (per day)
Sessions per day
Configuration - Puppet Hostgroups
Puppet Hostgroups
Databases - DBOD
DBOD Mysql DBOD Postgres DBOD Oracle
Databases - Transactions (per sec)
DB Transactions
Batch - Jobs (per day)
jobs
DFS - Disk (in TB)
TeraBytes
Used Total
EOS - Average Active Transfers (per hour)
Active Transfers
File Transfer - Average Throughput
Throughput
SLS Services - Availability (per week)
Empty Space

HEP Computing Overview for the next 10 years

- How could/should HEP computing evolve in the next decade?

- What is the role of CERN and IT Dep.?
  - How to address the needs of CERN
  - HL-LHC as focus, but other experiments too

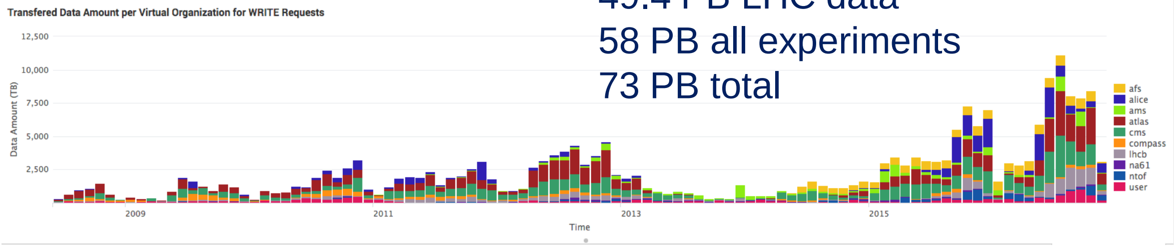

Evolution of Scientific Computing

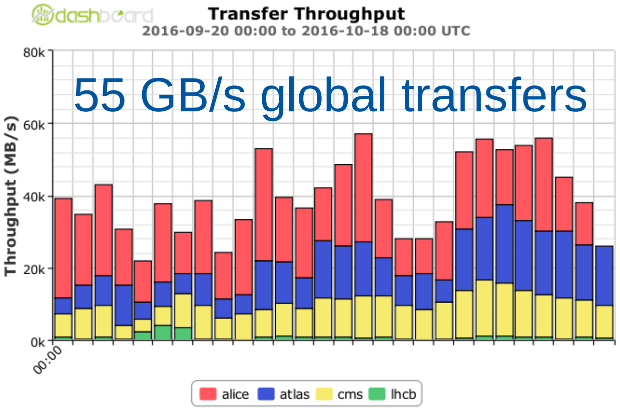
INTRODUCTION

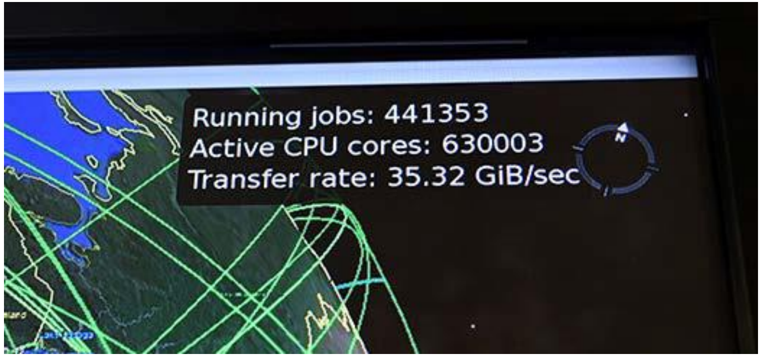
General Infrastructure

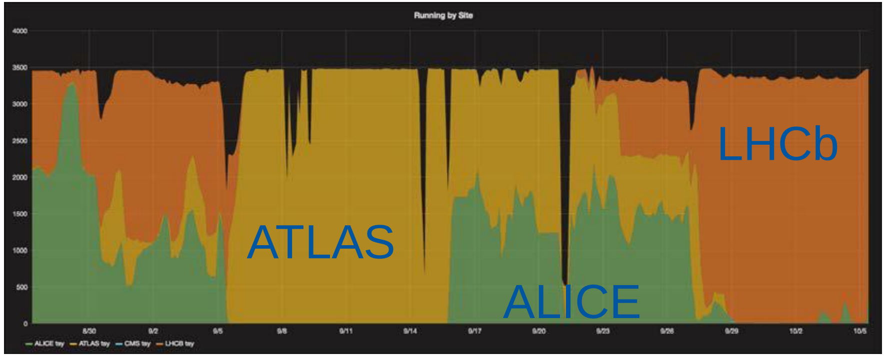
WLCG

2016:
49.4 PB LHC data
58 PB all experiments
73 PB total









LHC Run3 and Run4 Scale and Challenges

2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 …… 2030?

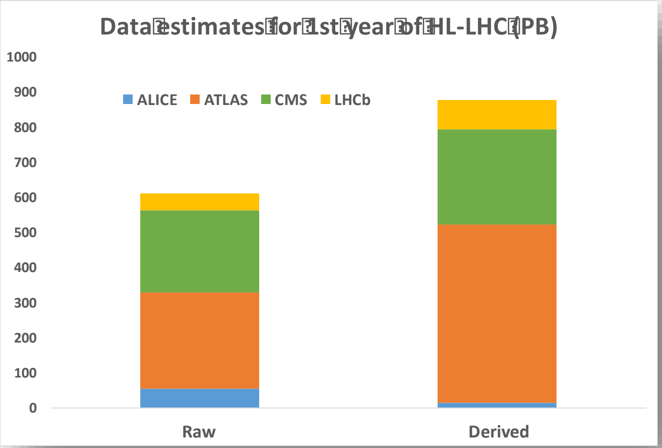
First run | LS1 | Second run | LS2 | Third run | LS3 | HL-LHC | FCC?

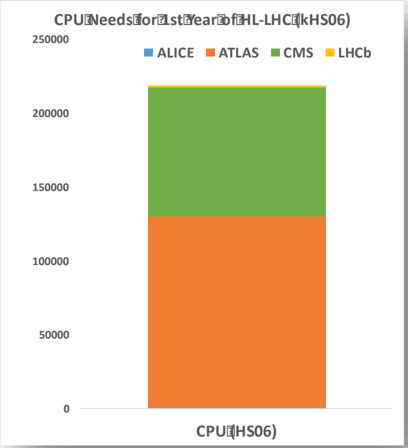
Raw data volume for LHC increases exponentially and with it processing and analysis load

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:

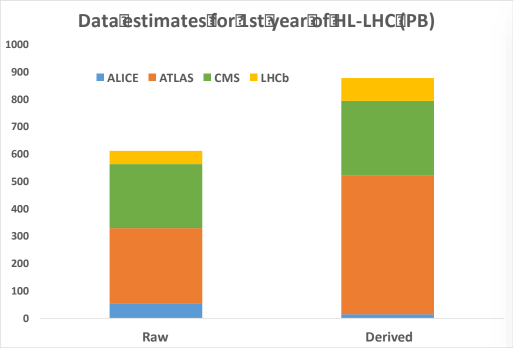
- Raw 2016: 50 PB → 2027: 600 PB
- Derived (1 copy): 2016: 80 PB → 2027: 900 PB

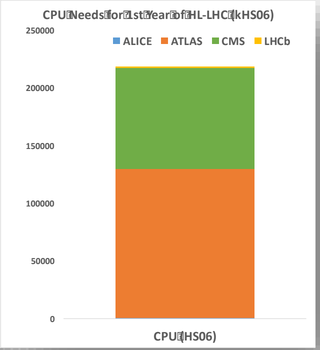
CPU:

- x60 from 2016

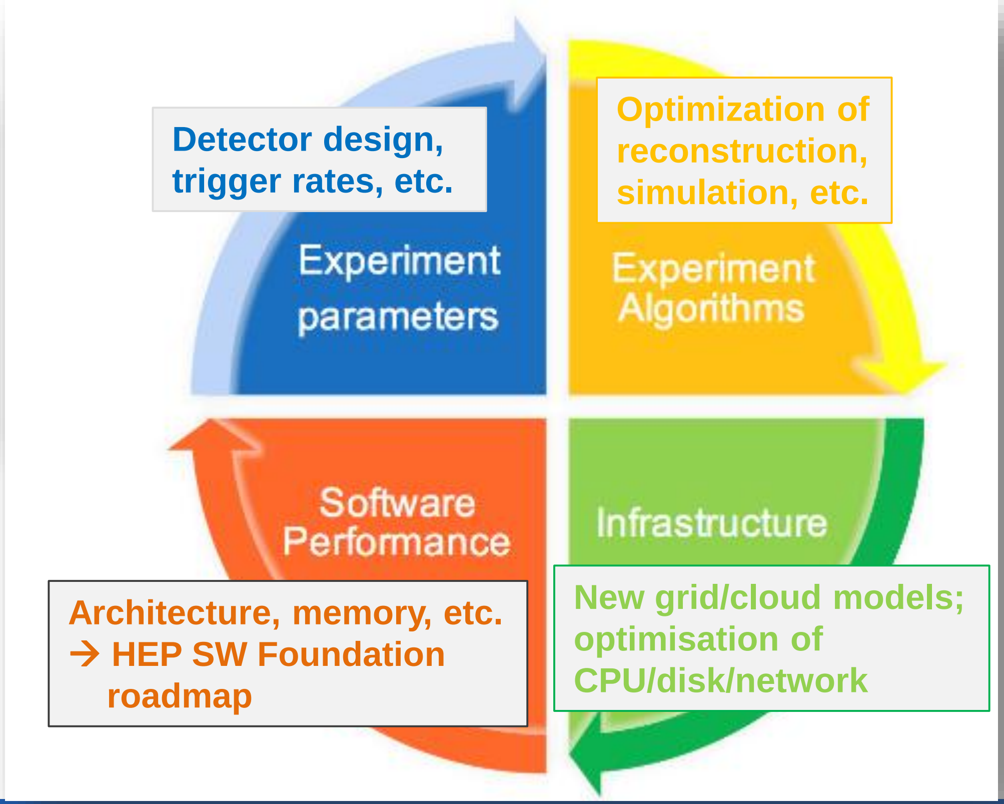
WLCG Achievements – 2

preparations for HL-LHC





HEP Software Foundation



1. CERN Services for LTDP

1. State-of-the art "**bit preservation**", implementing practices that conform to the ISO 16363 standard
2. "**Software preservation**" - a key challenge in HEP where the software stacks are both large and complex (and dynamic)
3. Analysis **capture and preservation**, corresponding to a set of agreed Use Cases
4. Access to **data behind physics publications** - the HEPData portal
5. An **Open Data portal** for released subsets of the (currently) LHC data **(later OPERA and others? …)**
6. A **DPHEP portal** that links also to data preservation efforts at other HEP institutes worldwide.

Update at DPHEP workshop in March 2017

iPRES 2016
13th International Conference on Digital Preservation //
Bern // October 3 – 6, 2016

2. What does DPHEP do?

- DPHEP has become **a Collaboration** with signatures from the main HEP laboratories and some funding agencies **worldwide**.
- It has established a "**2020 vision**", whereby:
  - All archived data – e.g. that described in DPHEP Blueprint, including LHC data – should be easily **findable** and fully usable by the **designated communities** with clear (Open) access policies and possibilities to annotate further;
  - Best practices, tools and services should be well run-in, **fully documented** and **sustainable**; built in common with other disciplines, based on standards;
  - There should be a DPHEP **portal**, through which data / tools accessed;
  - Clear **targets & metrics** to measure the above should be agreed between Funding Agencies, Service Providers and the Experiments.

First presented to ICFA in February 2013

2. How do we measure progress / success?

- **Practice:** through Open Data releases
  - Can the data really be (re-)used by the Designated Community(ies)?
  - What are the support costs?
  - Is this sustainable?
- **Theory:** by applying state of the art "preservation principles"
  - Measured through ISO 16363 (self-) certification and associated policies and strategies
  - Participation in relevant working & interest groups

One, without the other, is probably not enough. The two together should provide a pretty robust measurement…

2. ISO 16363 Certification

- Discussed already at 2015 PoW and several WLCG OB meetings
- Proposed approach:
  - An **Operational Circular** that describes the organisation's commitment to the preservation of scientific data & general principles (**draft exists**);
  - **Data Management Plans** by **project** where needed to refine embargo periods, designated communities etc.
  - A **Preservation Strategic Plan** covering a 3-5 year period
    - **DPHEP Blueprint (2012) and Status Report (2015) can be considered the first & second in such a series**
- This should cover the "**holes**" we have wrt section 3 of ISO 16363
- Needs to be done in close collaboration with experiments and other LTDP service providers: **start with a Workshop in 2017**
  - **Tentative dates: March 13 - 15 2017**

ISO 16363 metrics

Organisational Infrastructure (3)

3.1	Governance & Organisational Viability	Mission Statement, Preservation Policy, Implementation plan(s) etc.
3.2	Organisational Structure & Staffing	Duties, staffing, professional development etc. [APT etc.]
3.3	Procedural accountability & preservation policy framework	Designated communities, knowledge bases, policies & reviews, change management, transparency & accountability etc.
3.4	Financial sustainability	Business planning processes, financial practices and procedures etc
3.5	Contracts, licenses & liabilities, "rights"	For the digital materials preserved…

IT Challenges

Computing needs and technological challenges for the Data Center have been described this morning

The role of the IT Department is to keep providing an excellent level of services while scaling up to the foreseen resource levels at an affordable cost

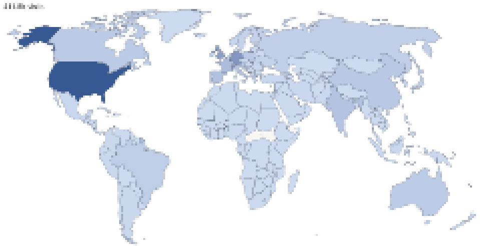
On-Premise Vs. Public Clouds

- Exclusive use of only on-premise or only cloud infrastructures not efficient, agile or cost-effective

- Hybrid computing and data infrastructures
  - CERN/HEP managed resources based on LHC needs
  - Transparent integration of commercially available resources for elastic scaling out as needed
  - Take advantage of Opportunistic Computing

Open Science: Zenodo

- Infrastructure
  - Runs on 30 VMs in CERN Cloud
  - 8TB of data storage (AFS > EOS)
  - Major SW upgrade in Sept (9 months prep)
  - [1 Staff (60% EC), 1 FELL, 1 TECH]
- Impact
  - Biggest issuer of DOIs for SW in world
  - Reference material for publications
    - F1000, Wiley, eLife, PLoS, Elsevier, Nature, etc
  - Recommended by EC and National programmes
- Pilot the “Cloud Credit” model
  - With NIH in their Commons
  - EC - metered usage recharging?
- Not just the long-tail
  - Frequent requests for 10TB-PBs of storage



Visitors from ~ all Countries
- Including
  - Antarctica
  - Vatican City
- 56% from Europe

57k Records
- 11k Software
- 3k Datasets

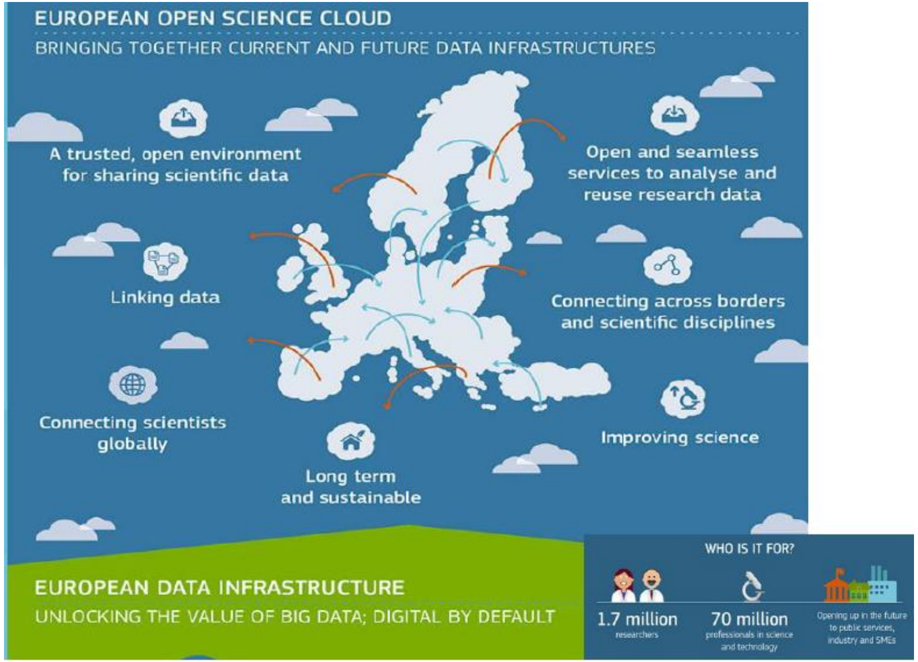
700 Communities
- Projects
- Institutes
- Subjects
- Conferences
- Publishers

Common Facilities

- Achieve economies of scale and operational efficiency by identifying additional common facilities across the LHC Experiments
  - LHCb/ALICE by 2021
  - ATLAS/CMS by 2025

- Investigation of possible models including the construction of a new (efficient) data center

International Collaborations



The objective of CERN's participation in the work programme is to develop policies, technologies and services that can support the Organization's scientific programme, promote open science and expand the impact of fundamental research on society and the economy.

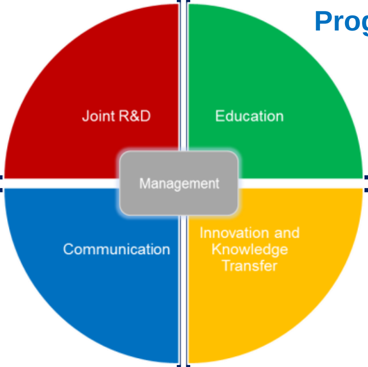
European Open Science Cloud will be the context for future projects

- We have to see the opportunities to contribute to it with CERN technologies and services
- We are positioning ourselves to give input to the new program 2018-2020

R&D and Innovation

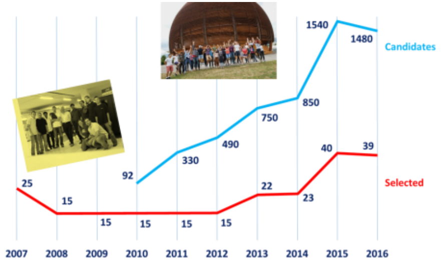
intel ORACLE SIEMENS IDT rackspace BROCADE
SEAGATE Yandex HUAWEI COMTRADE

CERN openlab



Seminars, training courses, academic training

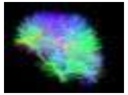
CERN openlab Summer Students Programme



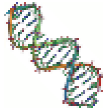
- 200+ press cuttings
- 150,000 visits to our website
- 50+ events, visits, lectures
- 100+ presentations
- 50+ news articles, press releases, case studies

15 years CERN openlab

Innovation & Entrepreneurship



BioDynaMo
The Biology Dynamic Modeller



GeneROOT

Applications to cross-disciplinary research

R&D and Innovation

Growing interest in the community for new models and tools (in addition to what will be described). For example:

- Applications of Machine Learning
  - What is the impact on the computing models and the data center resources?
- Applications of IoT-like infrastructures
  - Impact on network requirements? Security?

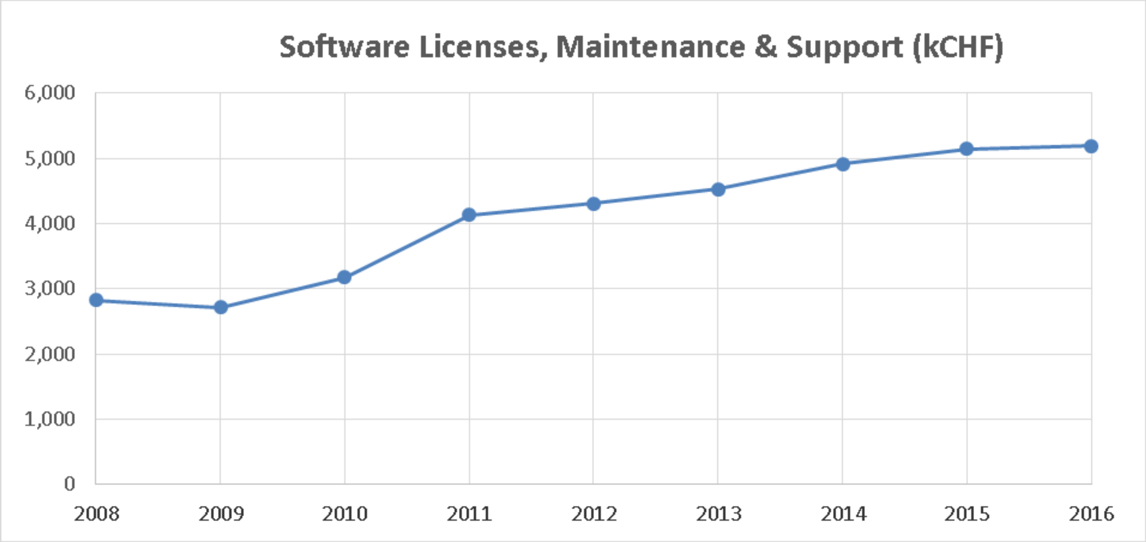
Education and Skills Development

Bi-directional requirements

- The introduction of new technologies requires new skills
  - Engagement in educational and training programs (e.g. Marie-Curie, like the ICE-DIP project; openlab summer student program; etc.)

- Ensure that young engineer/scientist leaving CERN have state-of-the-art skills readily usable in their future careers

Increasing Incompressible Expenses for Software Licenses, Maintenance and Support



- Risk to see software costs continue to increase in the near future due to commercial trends impacting the academic world.
- *Remark: License infringement by very few individuals has a significant cost*.