Joint Public/Private Research Projects

Alberto Di Meglio – CERN openlab Head

DRIVING INNOVATION SINCE 2001

Now in our sixth three-year phase

SET UP 2001

I 2003

II 2006

III 2009

IV 2012

V 2015

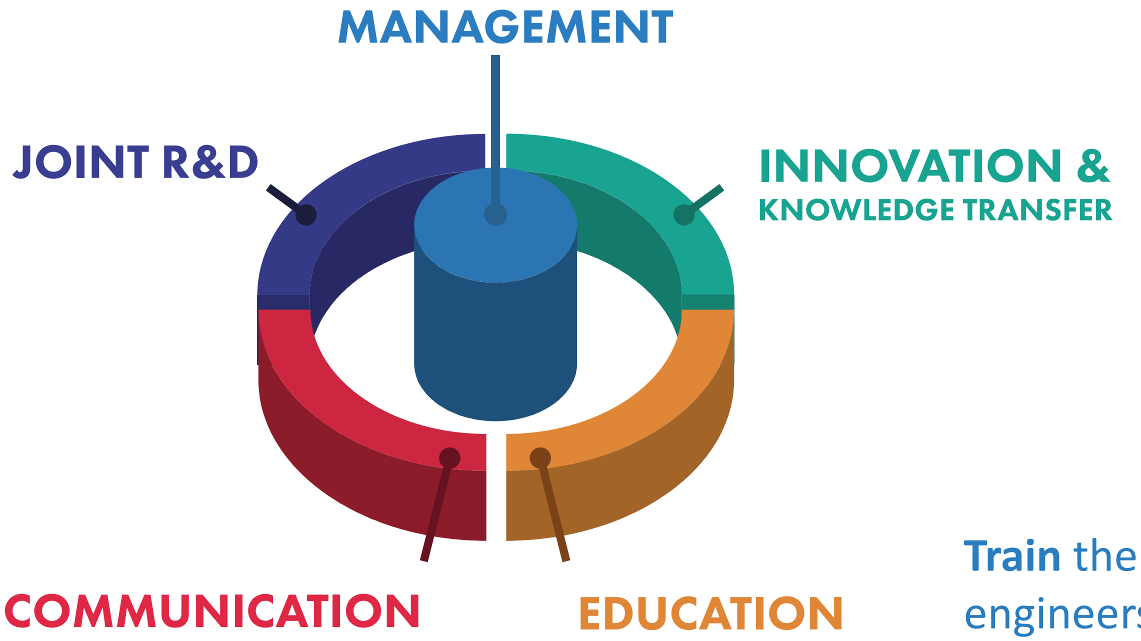
2018 Sixth Phase VI

CERN openlab Collaboration Board 2017

CERN OPENLAB'S MISSION

Our recipe for success

Evaluate and test state-of-the-art technologies in a challenging environment and improve them in collaboration with industry.



Collaborate and exchange ideas with other communities to create knowledge and innovation.

Communicate results, demostrate impact, and reach new audiences.

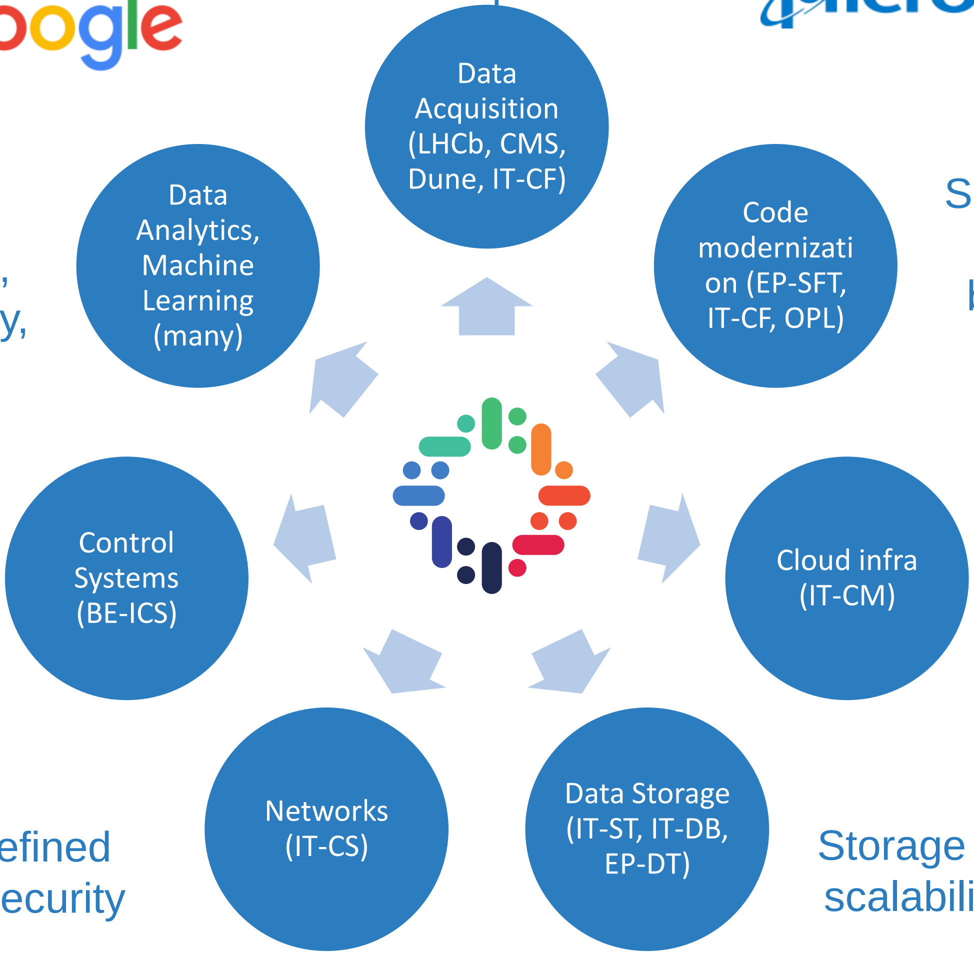
Train the next generation of engineers/researchers, **promote** education and cultural exchanges.

JOINT R&D PROJECTS (PHASE VI)

High-bandwidth fabrics, accelerated platforms for data acquisition

Fast simulation, Data quality monitoring, anomaly detection, physics data reduction, benchmarking/scalability, systems biology and large-scale multi-disciplinary platforms

Simulation, HPC on the Cloud, benchmarking



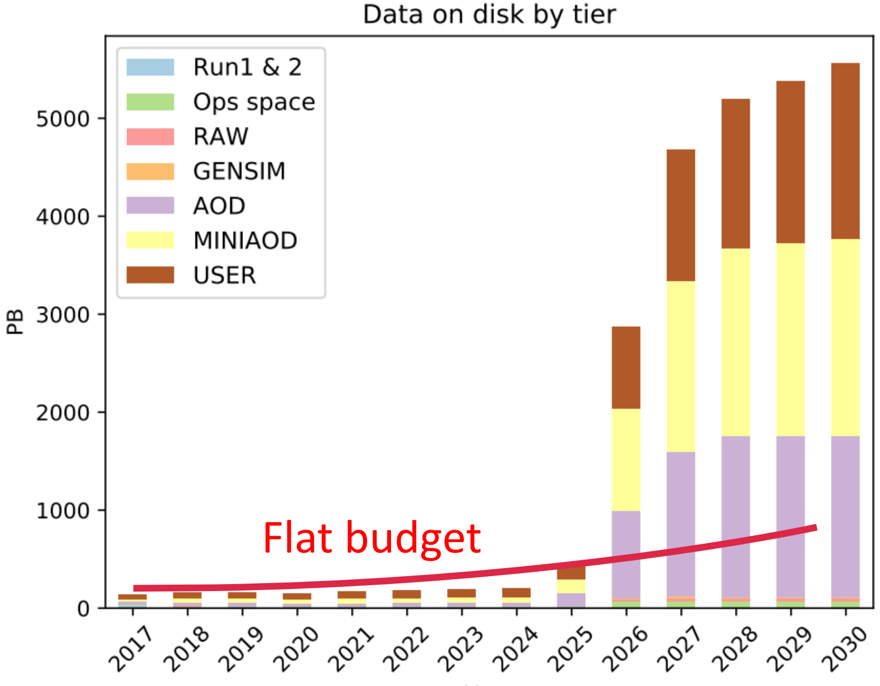
Predictive/proactive maintenance and operations

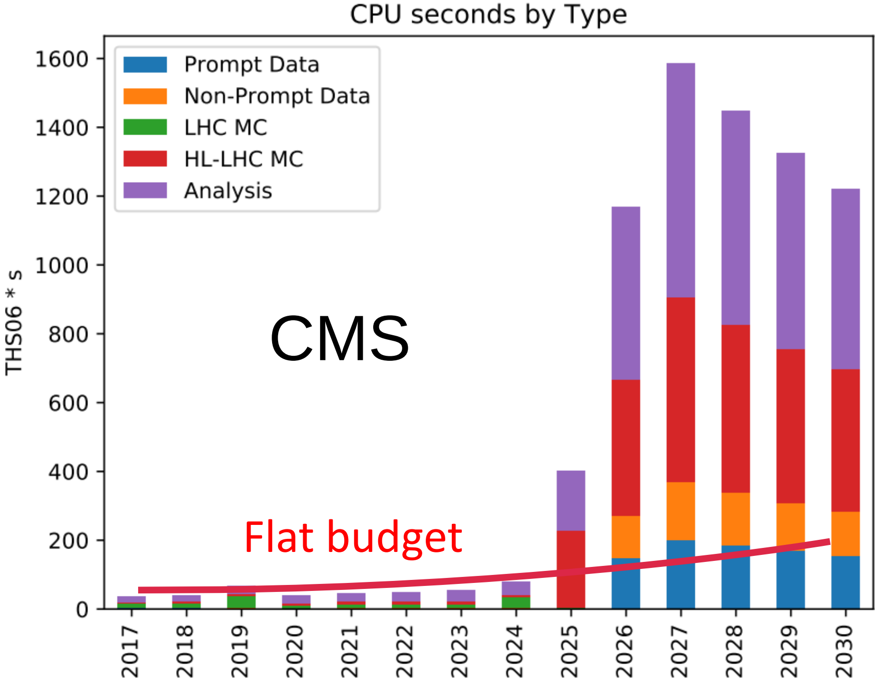
Cloud federations, containers, scalability

Software Defined Networks, Security

Storage architectures, scalability, monitoring

Computing @ HL-LHC





https://arxiv.org/pdf/1712.06982.pdf

Raw data volume increases exponentially

- Processing and analysis load

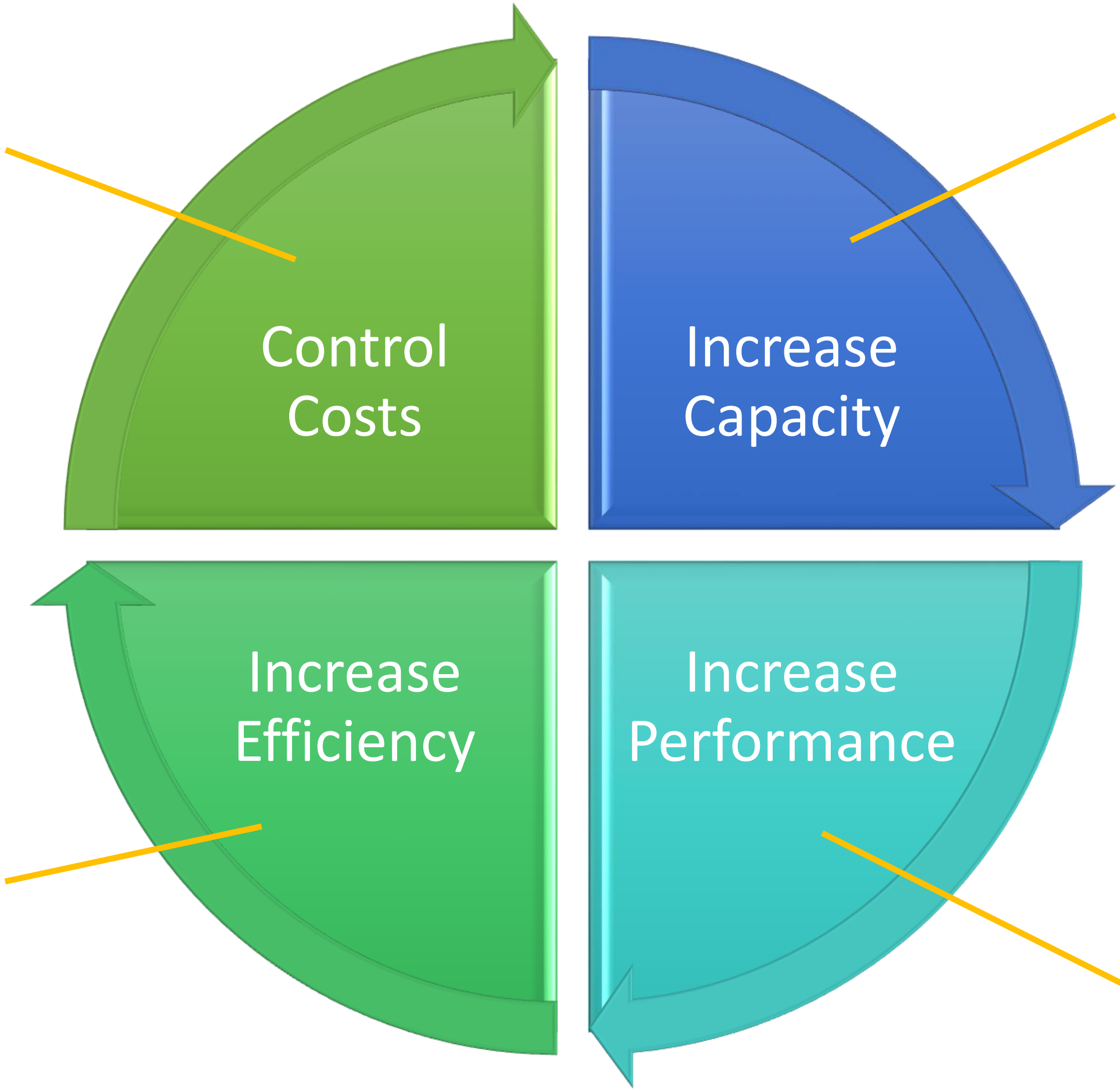
Technology at ~20%/year will bring x6-10 in ~10 years

- Estimates of resource needs x10 above what is realistic to expect

Computing and Data Challenges

Budget is not increasing

More computing and data processing/storage resources



Change the way processing is done

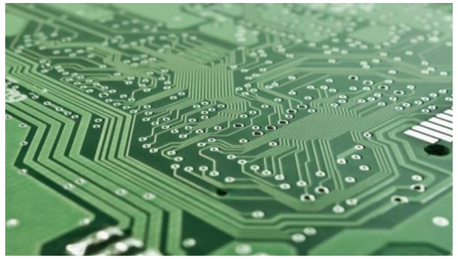
Faster Processing

Three Main Areas of R&D

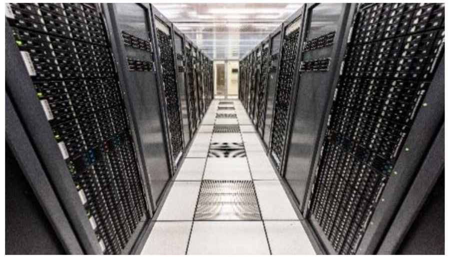
High Luminosity LHC

COMPUTING CHALLENGES

Increase **data centre performance** with hardware accelerators (FPGAs, GPUs, ..) optimized software

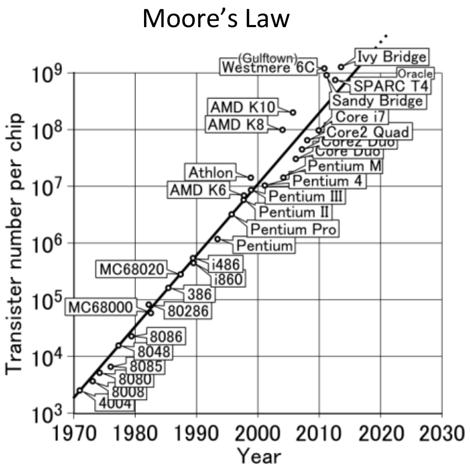


Scale out capacity with public clouds, HPC, new architectures



Change the computing paradigms with new technologies like **Machine Learning, Deep Learning**, Advanced Data Analytics, **Quantum Computing**

New Computing Platforms



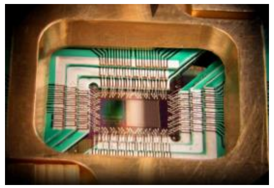
1995

2000

2010

2014

2015



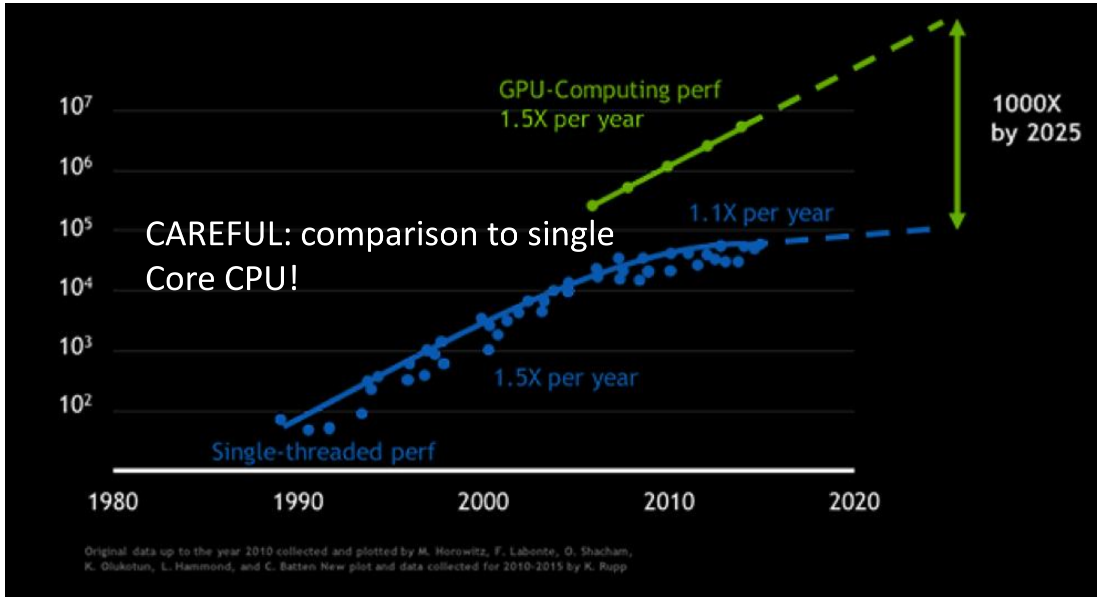
… and beyond

Radically new computing platforms are rapidly moving from pure computer science to realistic devices, e.g. **Neuromorphic Computing** and **Quantum Computing**

A **Quantum Computing Initiative** has been launched in December 2018 as a long-term investigation activity

GPU Acceleration

- Extremely parallel architectures are widely used in industry and other sciences
  - Can provide better performance/price and energy efficiency
  - GPU performance is on a an 18-month doubling cycle



- R&D effort in the LHC experiments to offload parts of the reconstruction workflows to GPUs
  - For both high level trigger and offline (on CUDA)

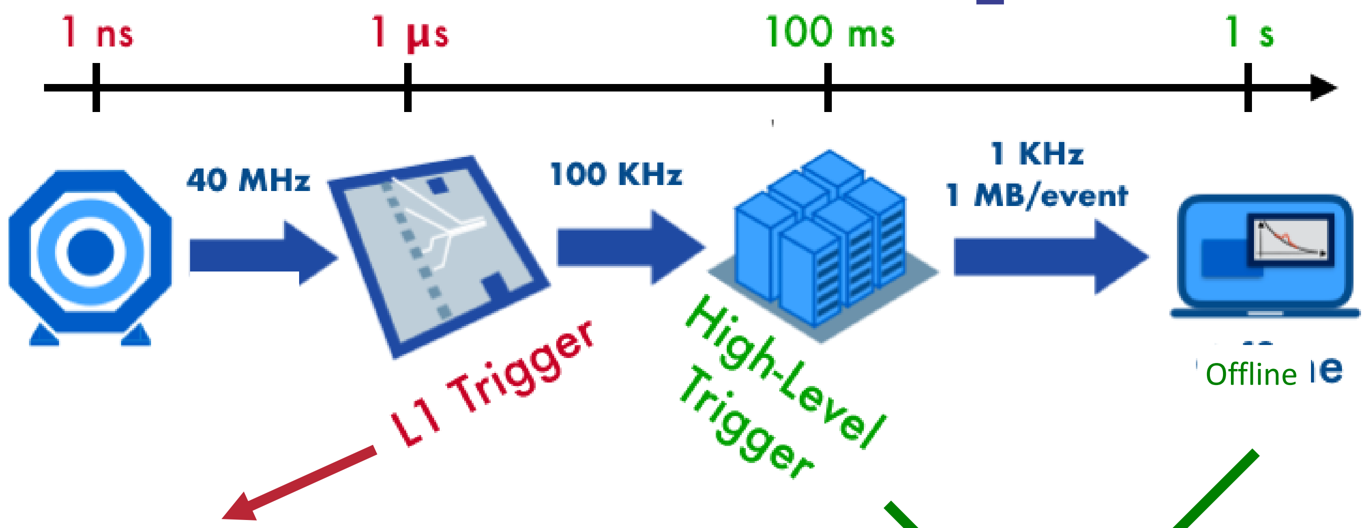
- ML/DL frameworks profit from GPU acceleration and substantially reduce turnaround times
  - Industry has invested heavily in optimizing machine learning libraries on GPUs (training and inference)
  - Increasing R&D effort in HEP

Industry is producing low-power GPUs

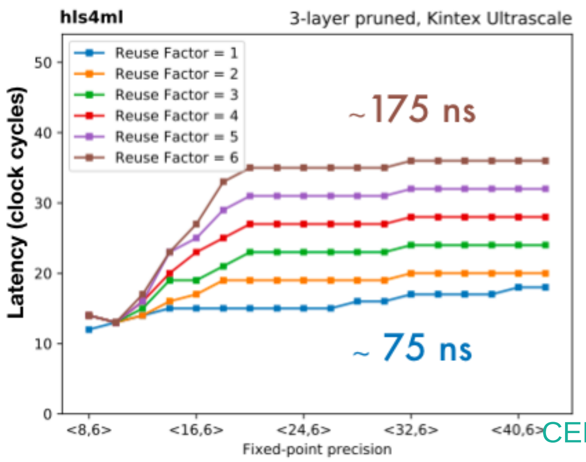
- Can be used to upgrade existing machines
- Latest **NVIDIA** GPUs (i.e. T4) use 25% of the power at significantly lower cost
- **AMD** GPUs offer very interesting performance/price ratios
  - Software is catching up

CERN openlab E4 COMPUTER ENGINEERING NVIDIA

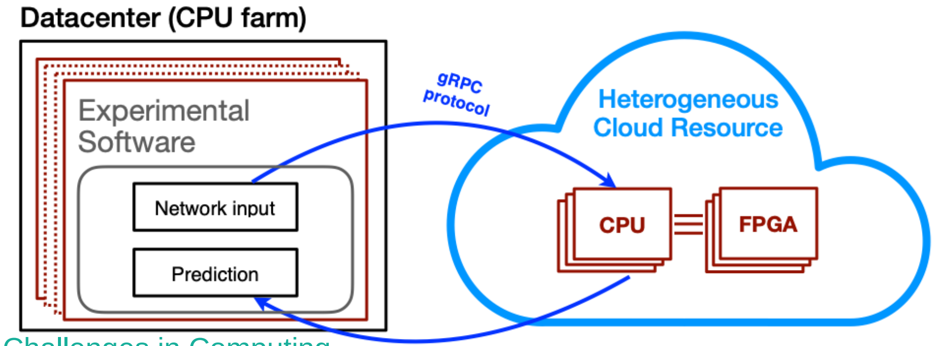
FPGA-accelerated Deep Learning



HLS4ML: Fast inference of deep neural networks in FPGAs for particle physics
(JINST, 13, 2018)



SONIC: FPGA-accelerated machine learning inference as a service for particle physics computing
(arxiv.1904.08986)



HEP Quantum Computing Initiative

- Launched in November 2018
  - Event at CERN > 400 participants
- Created a catalog of possible applications in HEP
- Established links with major QC activities in Europe and US

Currently discussing (among others) with

IBM

Google

intel

CQC CAMBRIDGE QUANTUM COMPUTING

IDQ FROM VISION TO TECHNOLOGY

Goal to establish a “hub” or “centre of excellence” in Quantum Computing applications

Research paths in Quantum Computing

- Tools and methodology development on emulators and simulators
  - Proof-of-concept algorithms for HEP workloads
  - Compare results on real devices
- Understand the role that CERN can play as part of broader QC development initiatives
  - APIs and user interfaces to access QC systems
  - Engineering aspects of QC installation (cryogenics, material science, ..

Examples

1. Quantum SVM for Higgs boson searches (University of Wisconsin)
   Higgs coupling to top quark
   Higgs decay to muons
   Double Higgs production
   Dark Matter
2. Quantum SVM for boosted Higgs (University of Washington)
   Identify jets originating from Higgs boson decay in order to study Higgs coupling to Standard Model particles
3. “Classical” optimisations
   Track reconstruction
   Grid workload optimisation

Storage Evolution

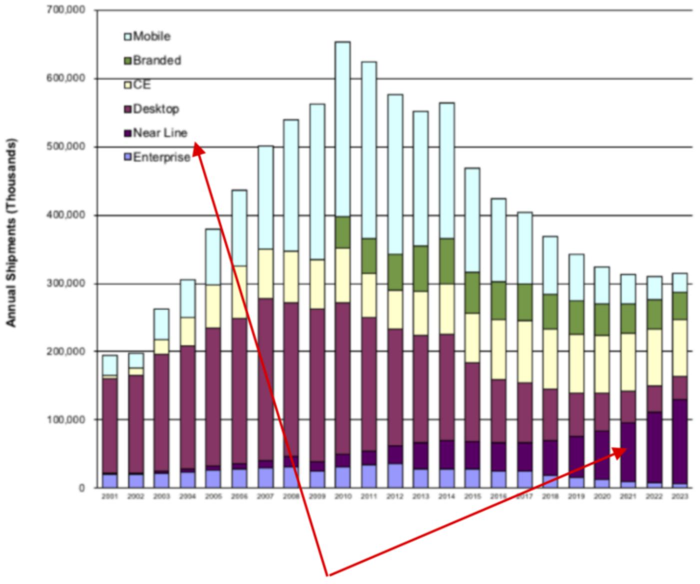
Storage remains a challenge as data volumes increase

Similar to computing, industry is specializing

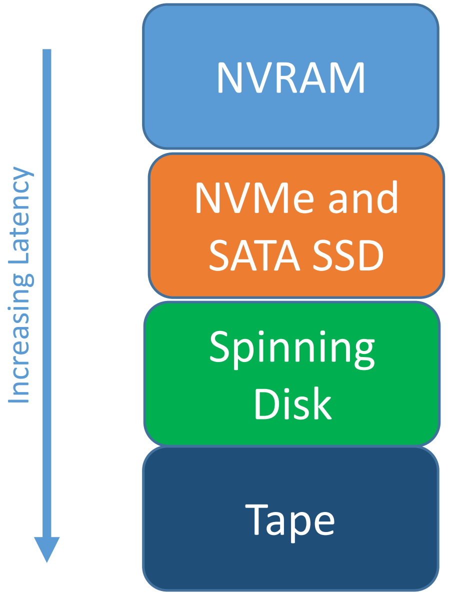
Improvements for HDD continue for capacity but slower for bandwidth and IOPS (Shingled disk, Helium drives, MAMR and HAMR). On-board processing (ARM/RISC V) might provide higher bandwidth, but still early to say

High capacity, fast and affordable SSD-flash based are taking over, thanks to the move from SAS/SATA to NVMe protocols

Smaller size, but extremely fast solutions like 3D XPoint NVRAM blur the boundary between memory and storage, but are still expensive



HDD type used for Cloud + HEP



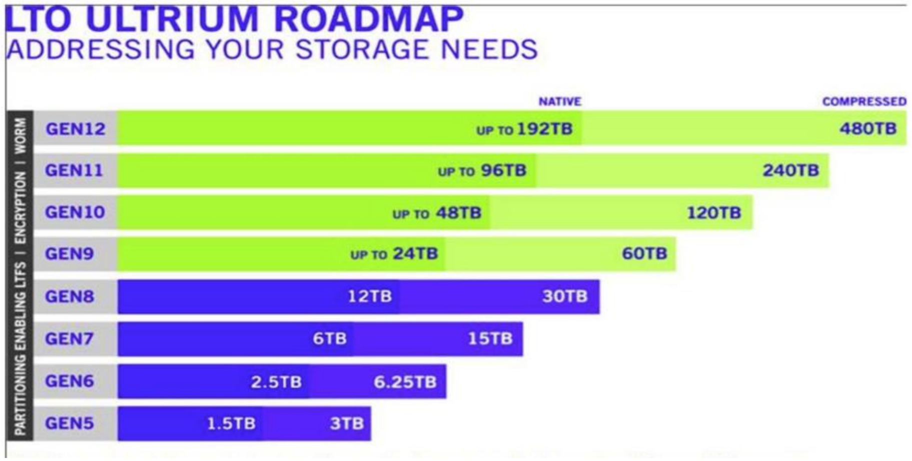
Archival Storage

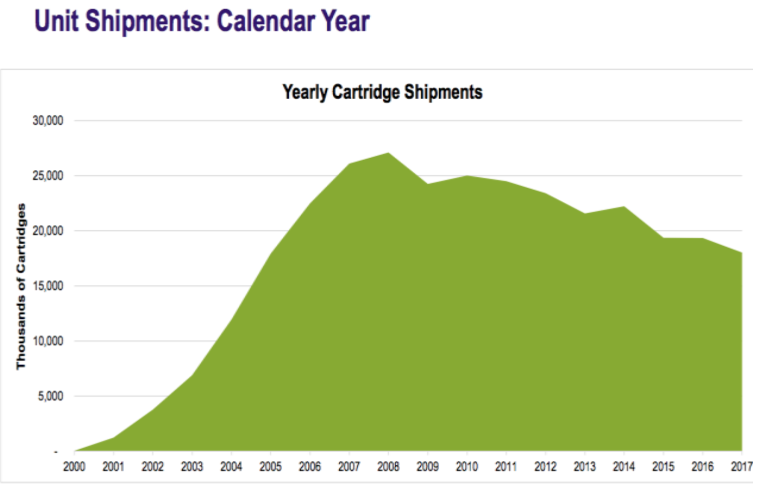
Tape remains the preferred (cheap) solution of HEP experiments for long-term storage

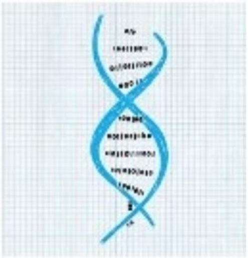
- As data volumes increase, LHC experiments are looking at larger roles for tape for data with very structured access
    - Tape roadmap extends another decade and a factor of 20 in capacity

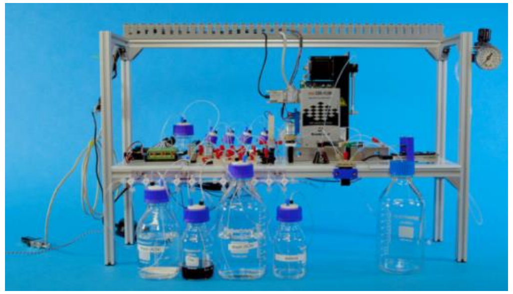
Disruptive technologies on the far horizon for archiving

- DNA Storage relying on the density and stability of DNA molecules to encode large volumes of data

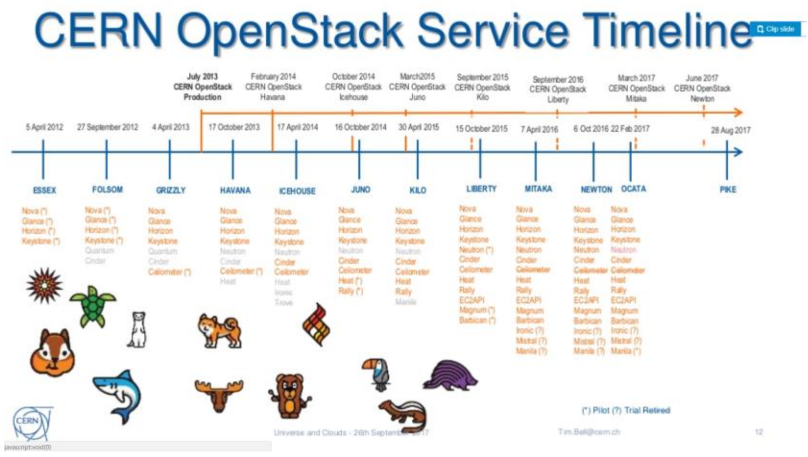


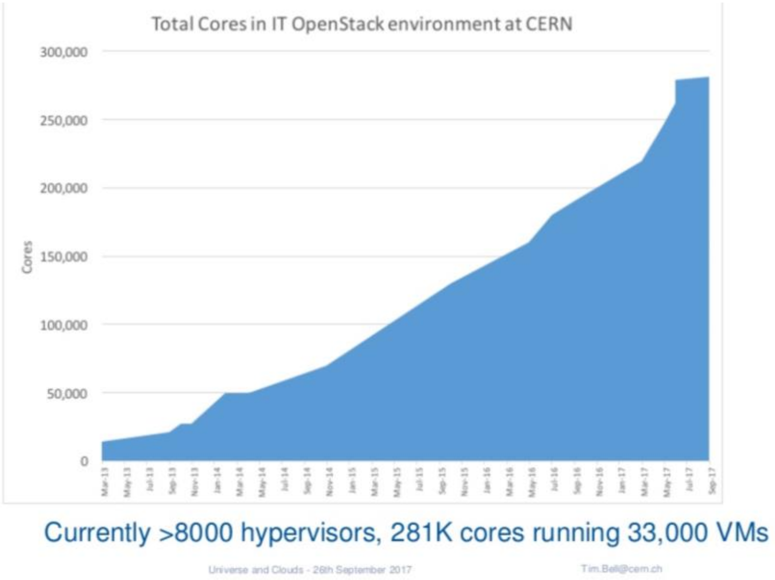






Cloud Computing





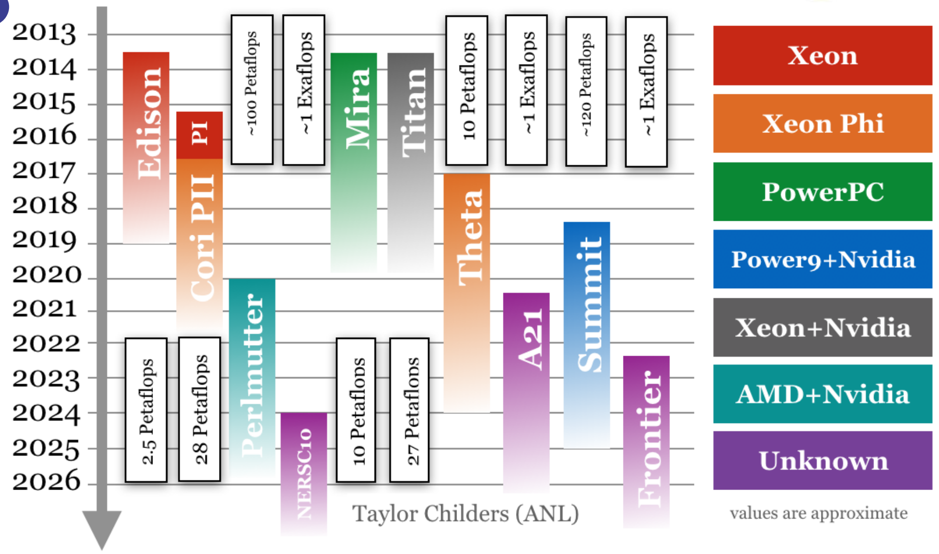
CERN runs the **largest OpenStack-based private cloud** in the scientific research world

Use of cloud technology allowed to **increase** the resource provisioning in **numbers** and **efficiency** by orders of magnitude

However, it is still limited by the maximum physical capacity of the DC

Hybrid public-private models are being considered, also to provision types of resources that we do not have (large GPU farms and dedicated data training facilities)

Investments at HPC Sites



HPC sites will grow by a factor 20 on the time scale of HL-LHC

- Large investments in USA, Europe and Asia
- Mostly available to sciences through allocations

High performance accelerated hardware will provide the bulk of the processing capacity (~10% expected from CPUs) . Implies:

- R&D on software and techniques to better exploit processing architectures, low latency networking, edge services, ..
- R&D in interoperability across HPC (and Cloud) sites, via consistent services (containers , CVMFS, ..)

Name	Supplier & Model	Compute power (Peak) [GFLOP/s]	Energy consumption under Linpack (kw)
Piz Daint gpu	Cray XC50 (Hybrid)	25'326'000	2'272
Piz Daint mc	Cray XC40 (Multicore)	2'193'000	734
Installed 2017			

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Engaging and working together with Super Computing Centres is essential for HEP and opens questions about training needs in software engineering

PRACE EuroHPC Joint Undertaking BDEC2 BIG DATA AND EXTREME-SCALE COMPUTING 2 DEEP Projects ETP 4 HPC EUROPEAN TECHNOLOGY PLATFORM FOR HIGH PERFORMANCE COMPUTING

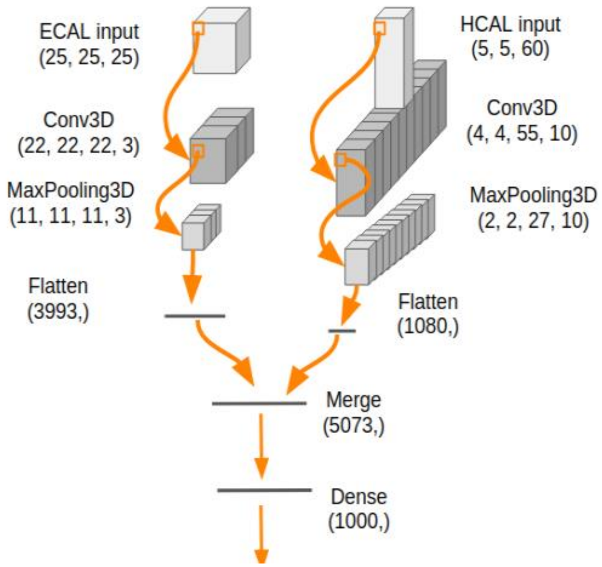
Courtesy of Maria Girone – CERN openlab

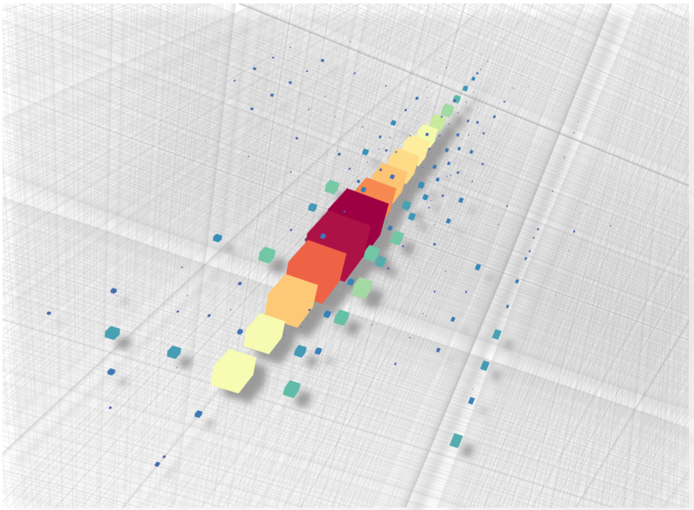
HPC Joint Activities

- **Active effort** in CERN openlab to support WLCG exploration of **common challenges in HPC integration** from individual LHC experiments documents
  - Draft document available at https://docs.google.com/document/d/1AN1d6NukhBsKnNH1MVvszqWdpcMfFGaYQMEVnS01Tc/
- Discussions with CSCS, Jülich, Oak Ridge to establish **pilot projects**
- Discussions with PRACE (workshop in October 2018)
  - https://indico.cern.ch/event/760705/
  - Proposal of MoU between SKA, CERN and GÉANT under evaluation
- **Leadership of CERN tasks** in EC-funded project **DEEP-EST** (Modular Supercomputing prototype at Julich)
  - 1 FTE for ECAL and HCAL reconstruction on GPU

Courtesy of Maria Girone – CERN openlab

Artificial Intelligence





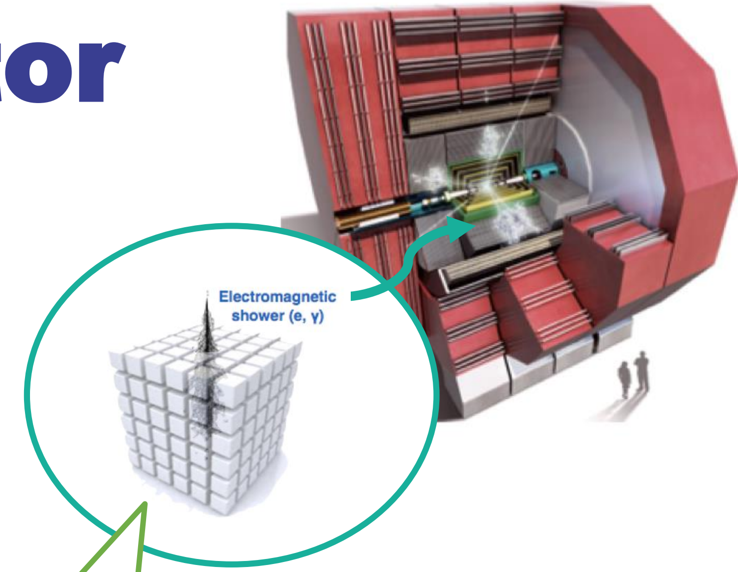
LHC Experiments have used **Machine Learning** for quite some time now (e.g. BDT), now investigating **Deep Learning** in several areas

New DL algorithms and **fast hardware** can accelerate almost all the typical HEP workloads (data acquisition, event classification, reconstruction, simulation, etc.)

Many ongoing projects in collaboration with Intel, Google, IBM, T-Systems, and others on algorithms (fast simulation, GANs, anomaly detection), tools (Spark, TensorFlow) and hardware (GPU, TPU, FPGA)

Deep Learning for fast detector simulation

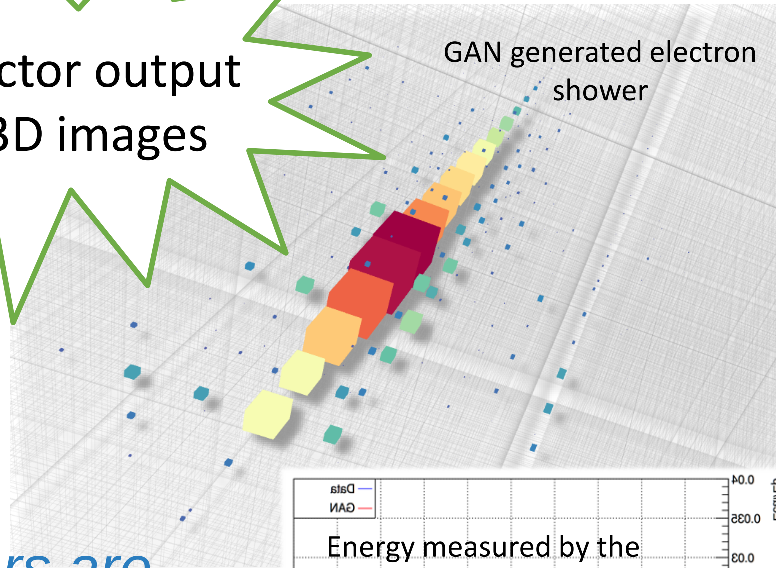
Detailed simulation of subatomic particles interactions is essential



Monte Carlo approach is not fast enough for HL-LHC needs

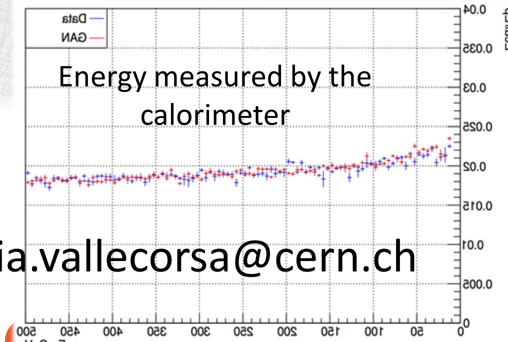
- **3D convolutional GAN** generate realistic detector output
- >2000x faster than Monte Carlo on a Intel Xeon processor

Detector output as 3D images



Training takes ~1 day on NVIDIA P100

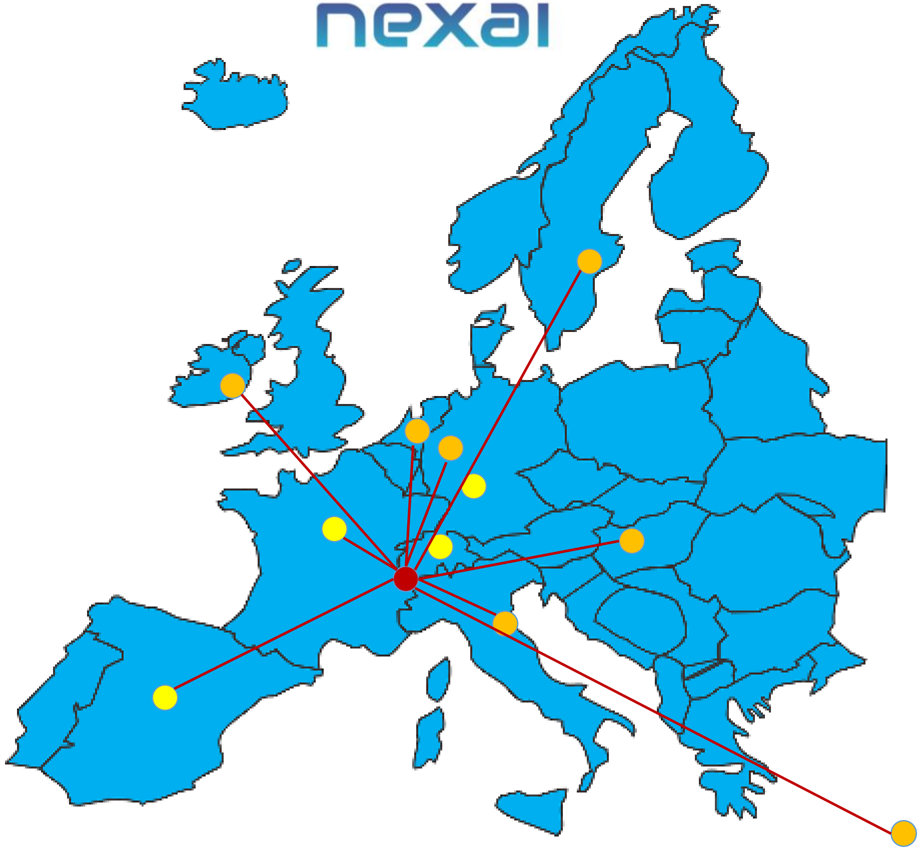
- Use parallel approach to distribute training across multiple nodes



*To generalise systems with larger memory and compute clusters are needed → **HPC workload***

Sofia.vallecorsa@cern.ch

Centre of Excellence in AI



Currently leading the preparation of a proposal for the creation of a Centre of Excellence in Artificial Intelligence

EC call ICT-48-2020

Tier 1-2 HPC Centres
Leading academic Institutes in AI research
Four applications domains:

- Fundamental physics research
- Personalised Medicine
- Sustainable Agriculture/Green Economy
- Dematerialized Banking and Security

Medical Research

Image Analysis

Machine-learning-based content recognition, offline and (quasi) real-time anatomy localization, tissue segmentation,
support to diagnostics,
automated/hybrid surgical rooms

Clinical Trials

Optimization, targeted recruitment and enrollment via automated analysis of medical records, pipeline assessment/review

Clinical Decision Support Systems

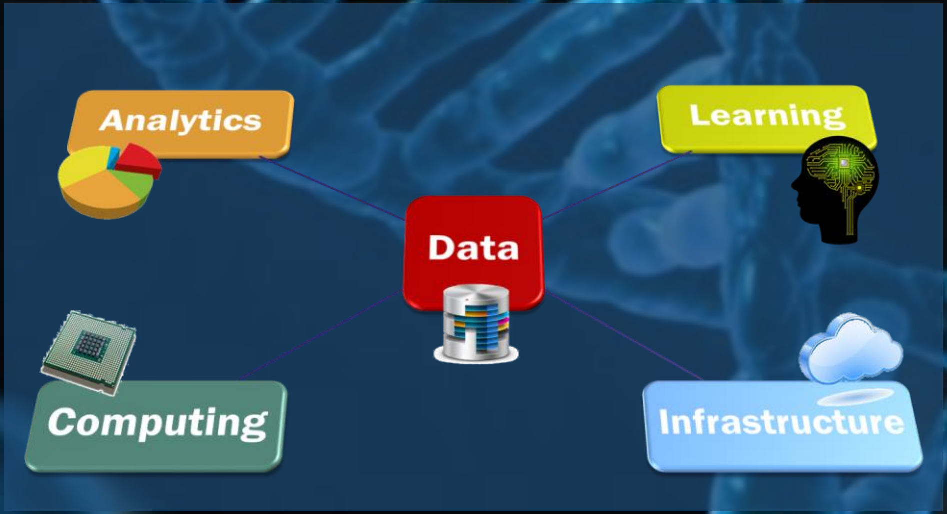
Large-scale data integration/analysis, structured/unstructured data, anonymization, treatment effectivess and risks predictions

Drug discovery

Combination of computing, simulations, textual descriptions of symptoms and expected outcomes

Trends and simulations

Preventive treatments, epidemics assessment and control, optimization of effort/budget allocations



Genomics

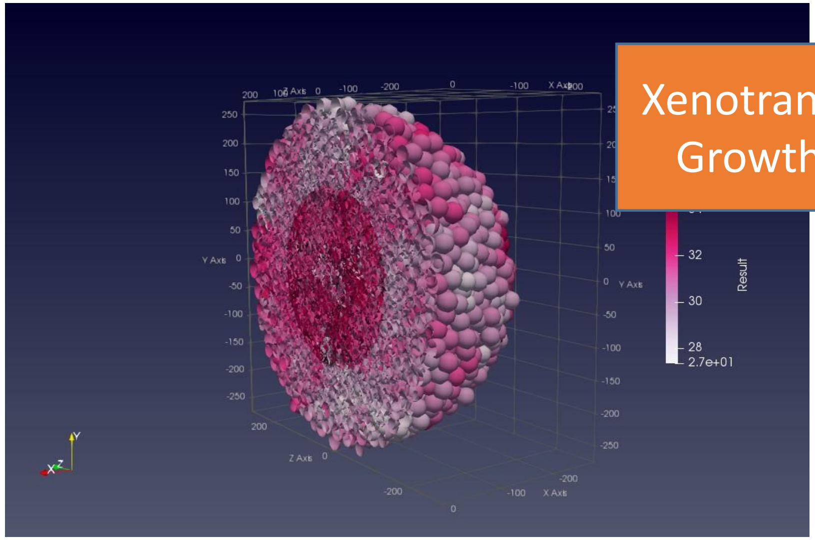
NGS data,
analysis, storage, transfer at multi-PB scale, integration with other *omics and clinical data, personalized treatments

Personalised Medicine

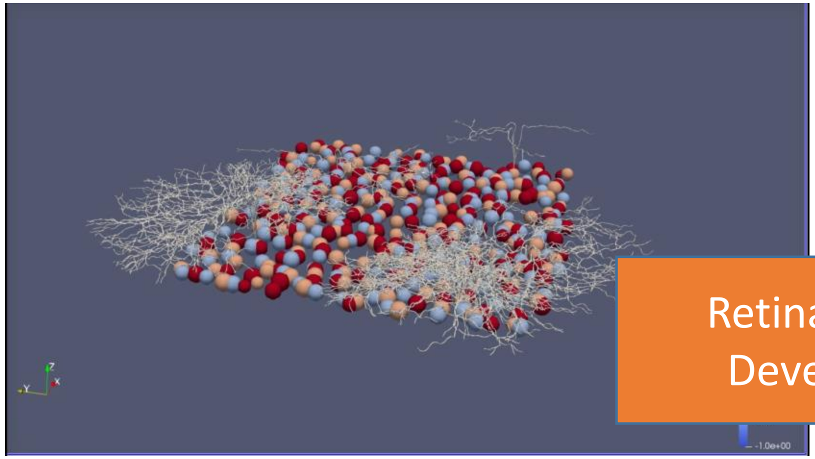
Monitoring/assessment, suggestions, personalized medicine, exploitation of data from wearable devices, social fitness/health platforms

Biological Dynamics:

BioDynaMo
Biology Dynamics Modeller



Xenotransplant Tumor Growth Simulation



Retinal Mosaics Development

Collaborative platform to support biological simulation experiments from single computers to large-scal cloud-based infrastructures

Extensible modular architecture able to support different cell types definitions, behaviours and interactions

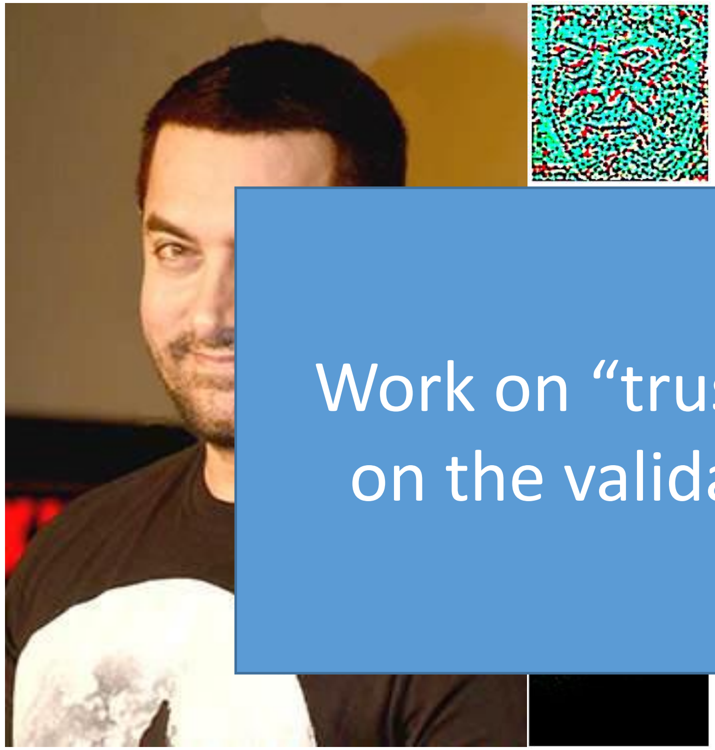
Growing library of tools, cell definitions, simulation workflows and models

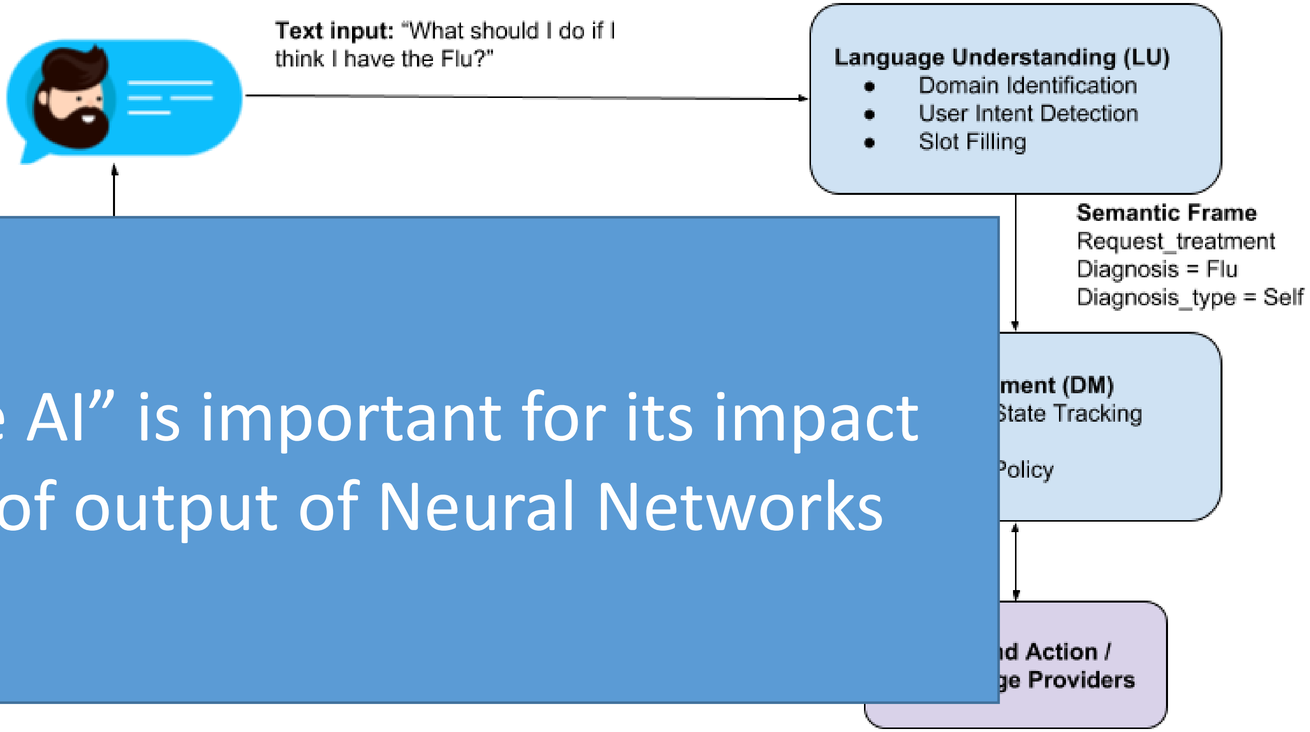
Potentially applicable to many fields, from developmental studies to in-silico treatments investigations to environmental application (sewage, waste treatment, etc.)

Open source license and contribution model

Machine Assisted Diagnostics

© Taghi Aliyev – CERN openlab - 2019





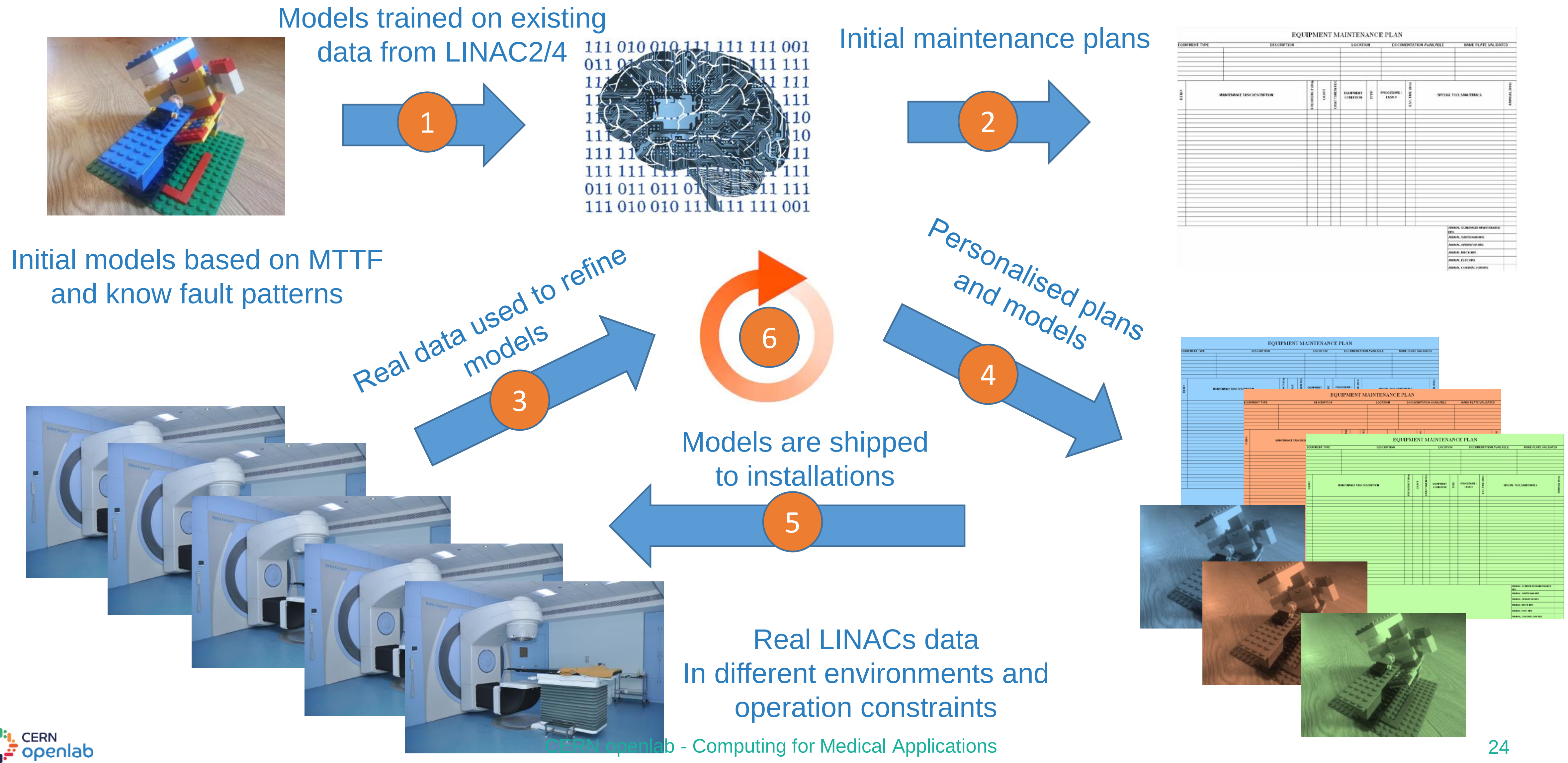
Young, 2000 - https://doi.org/10.1098/rsta.2000.0593

Work on "trustable AI" is important for its impact on the validation of output of Neural Networks

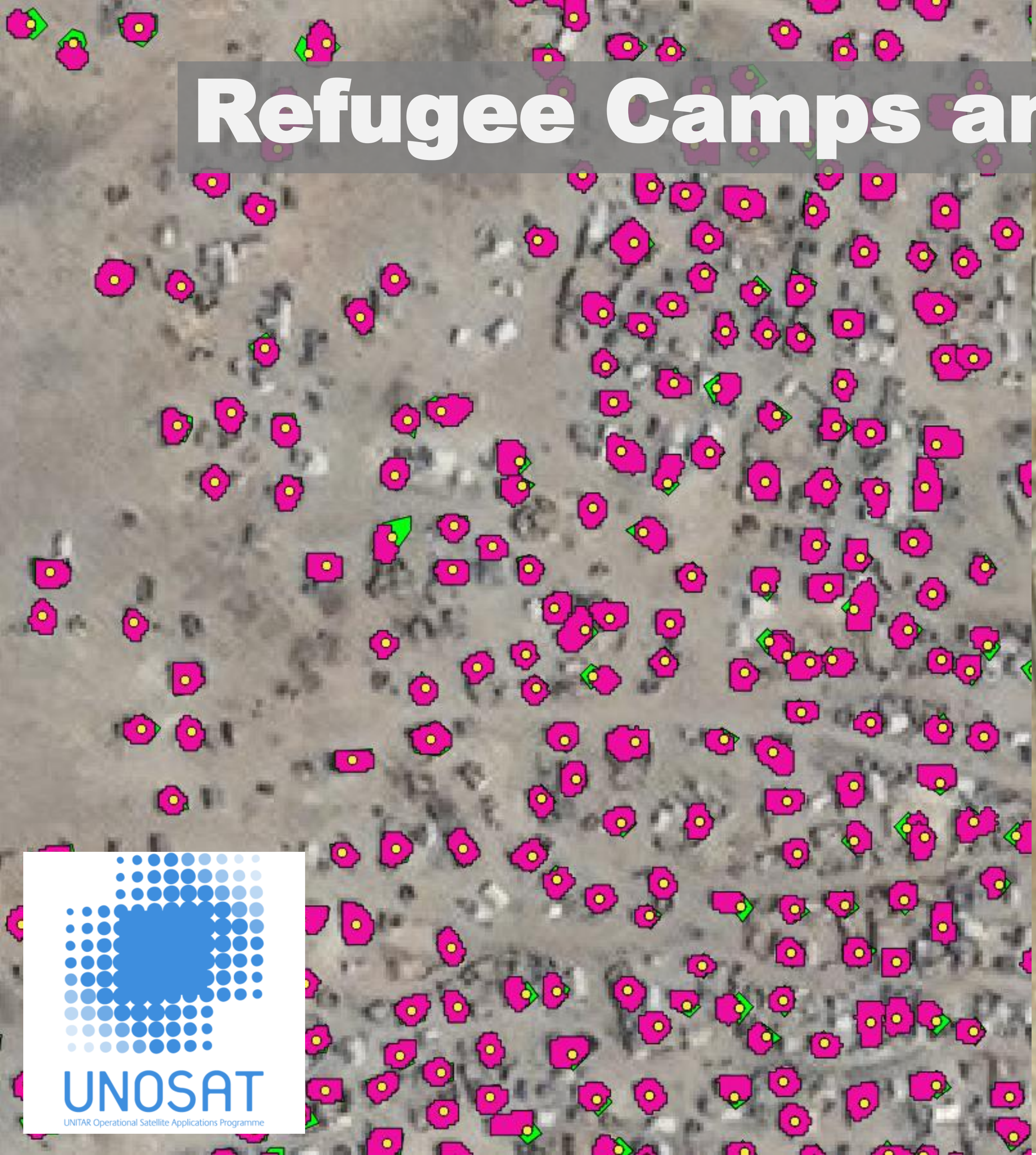
Phenotype Modelling and Genetic Evidence

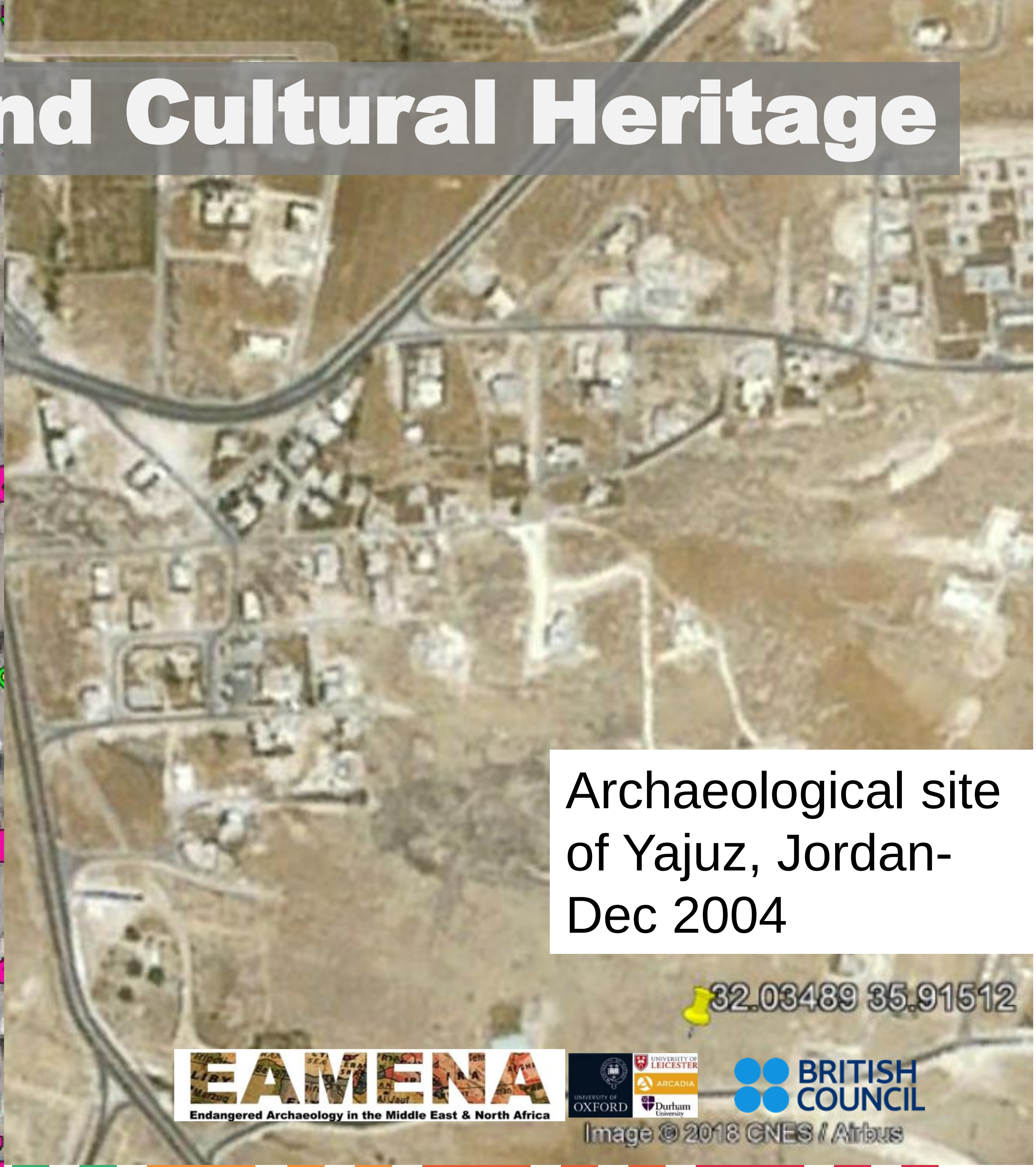
Conversational medical chatbots and Natural Language Processing

LINAC Personalised Preventive Maintenance



Refugee Camps and Cultural Heritage





Archaeological site of Yajuz, Jordan- Dec 2004

UNOSAT
UNITAR Operational Satellite Applications Programme

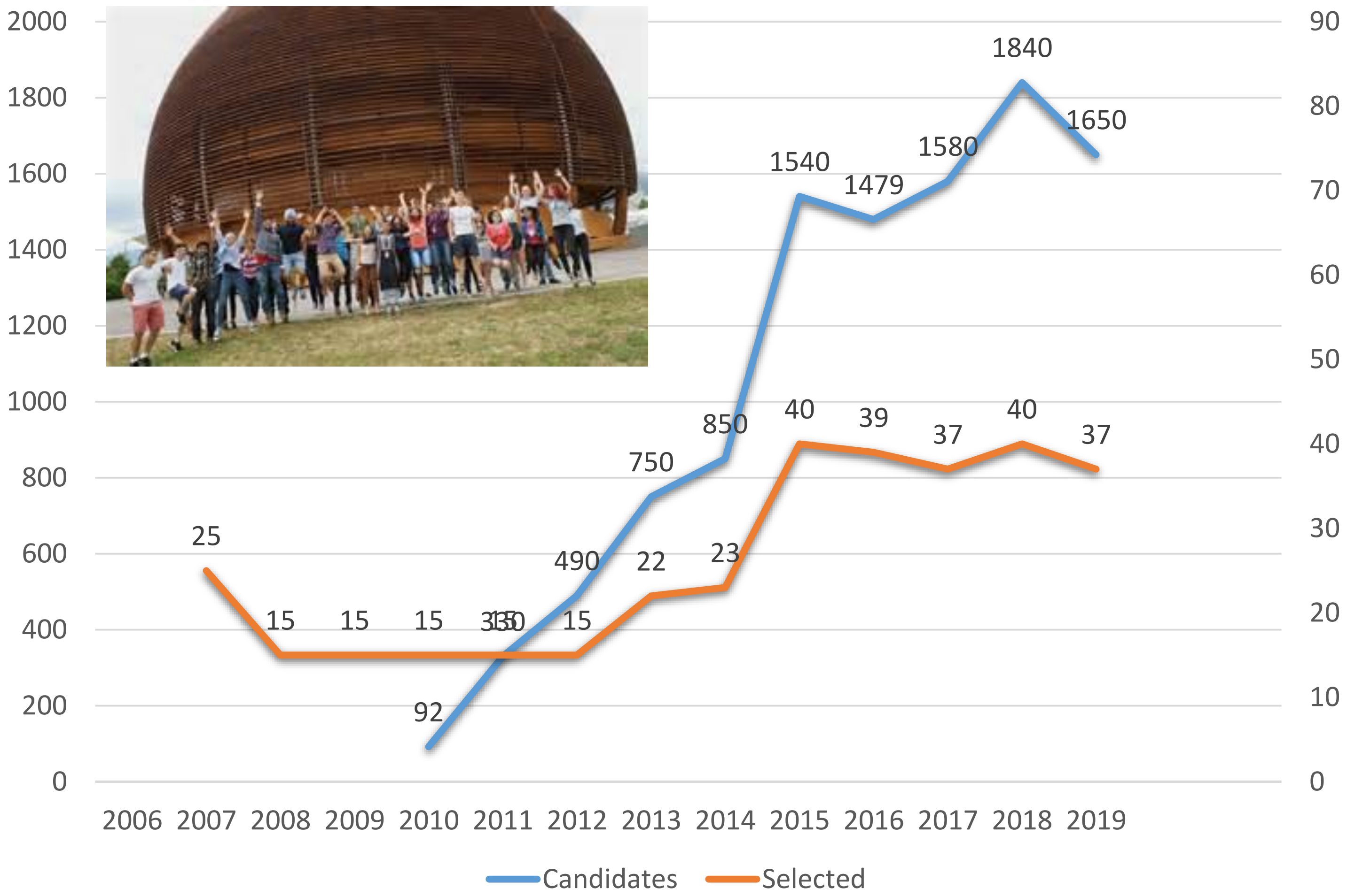
EAMENA
Endangered Archaeology in the Middle East & North Africa

University of Oxford
University of Leicester
Arcadia
Durham University

BRITISH COUNCIL

Education and Training

SUMMER STUDENT PROGRAMME



In 2019

- 1650 applicants
- 37 selected students
- 14 lectures
- Visits to external labs and companies
- Lightning talks session
- Technical reports

Thanks!

alberto.di.meglio@cern.ch
@AlbertoDiMeglio