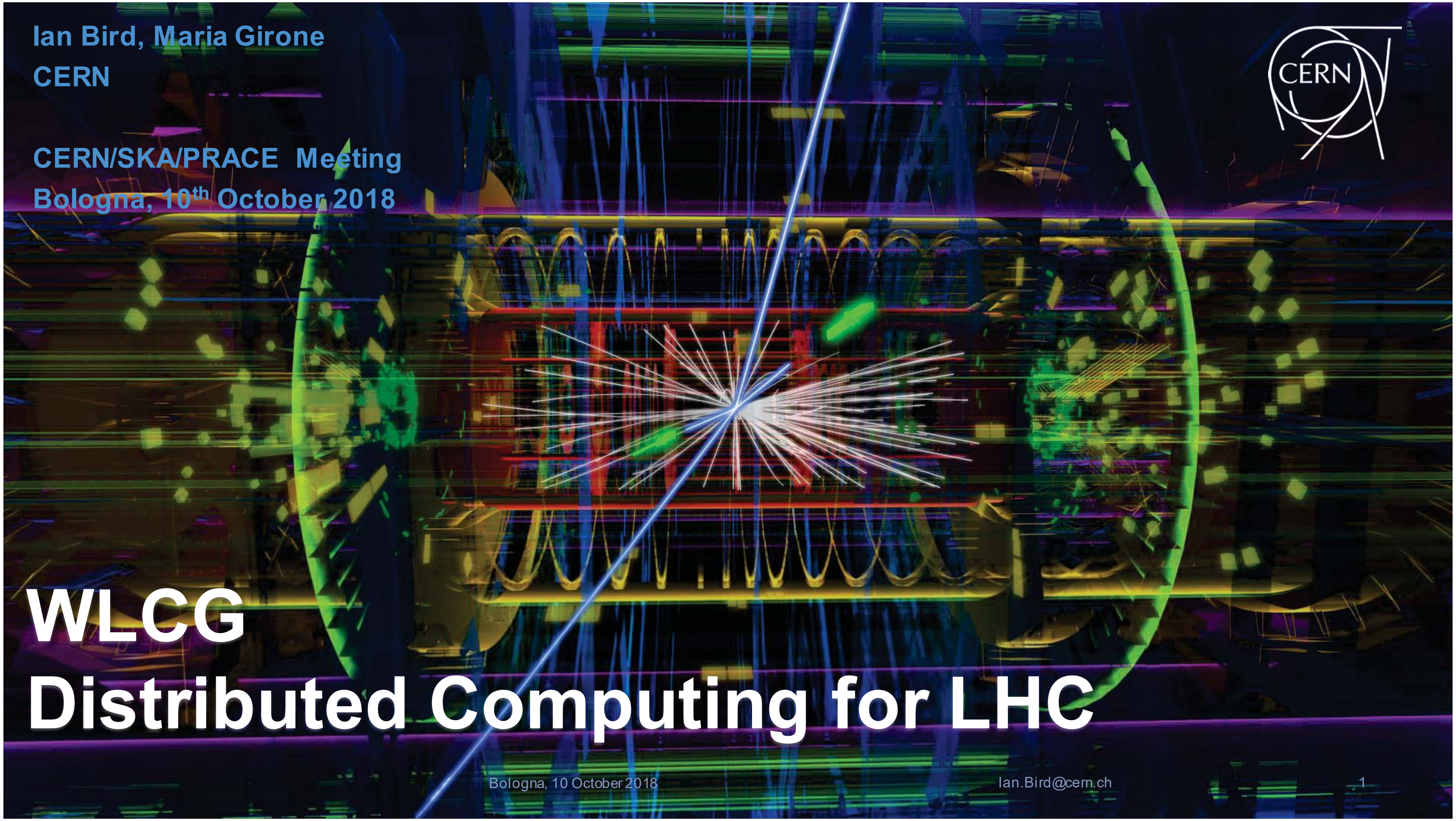


Ian Bird, Maria Girone
CERN

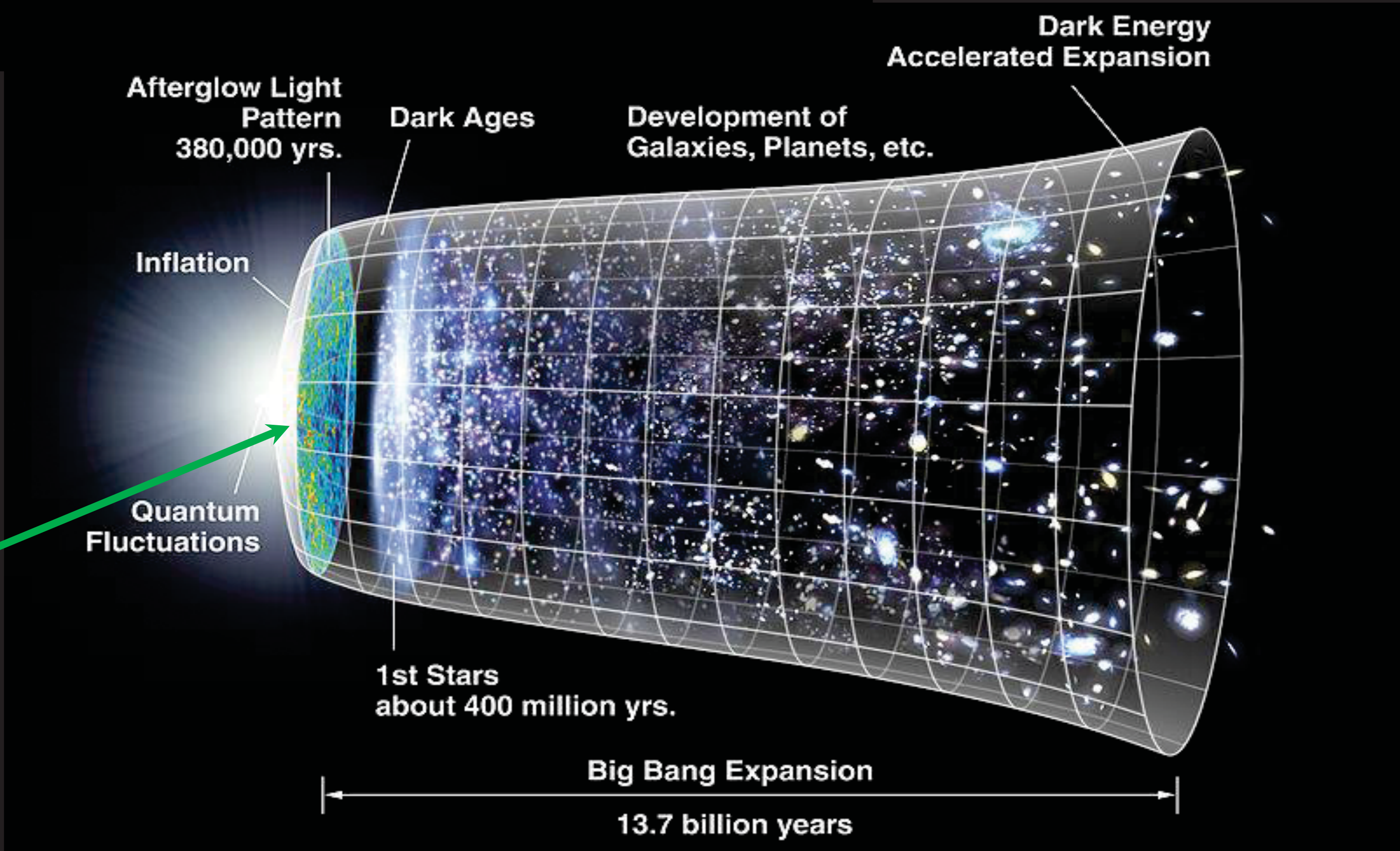
CERN/SKA/PRACE Meeting
Bologna, 10th October 2018

WLCG
Distributed Computing for LHC

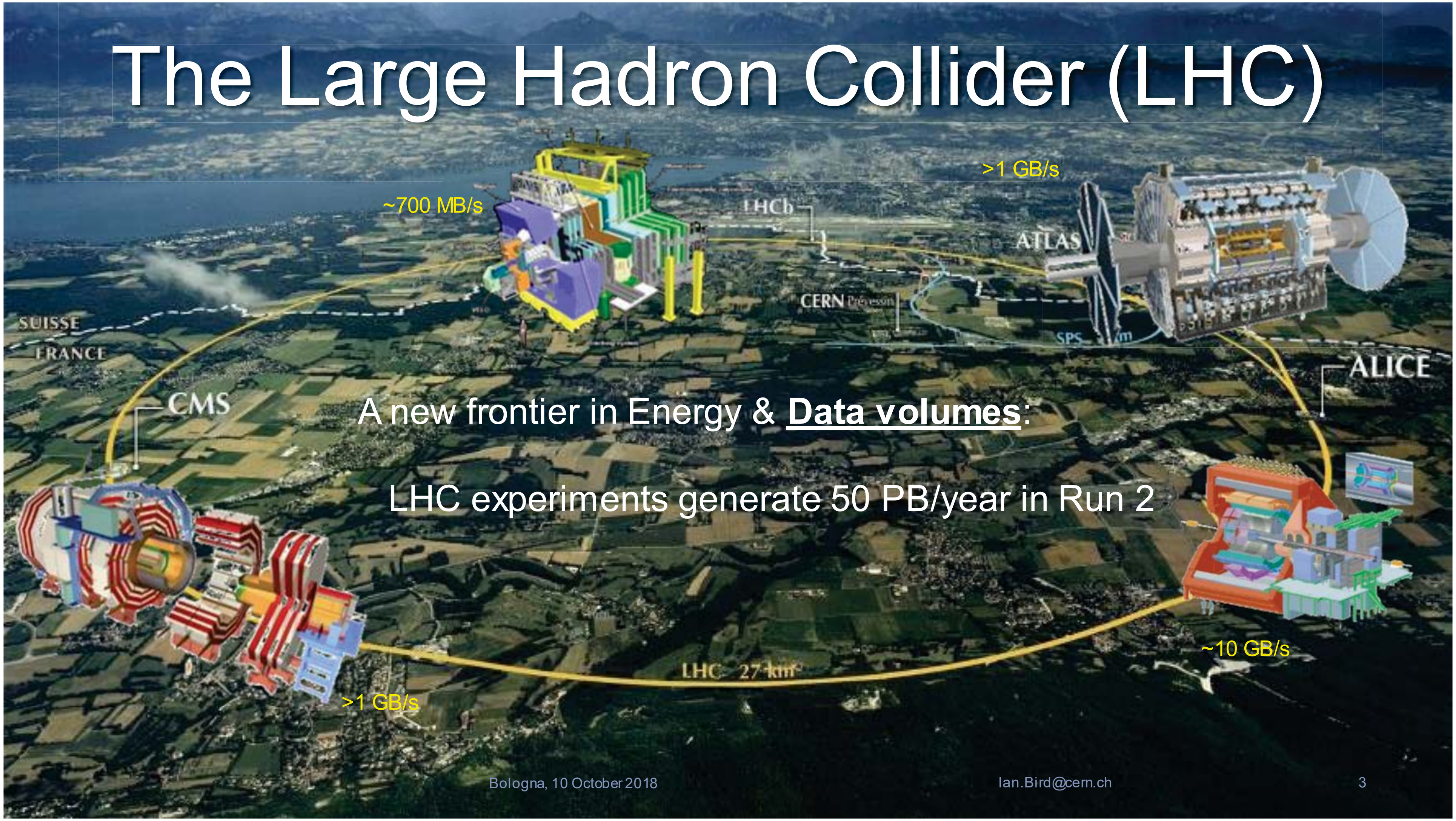
Evolution of the Universe

LHC drivers

Test the Standard Model
Dark Matter
Dark Energy
Anti-matter
(Gravity?)

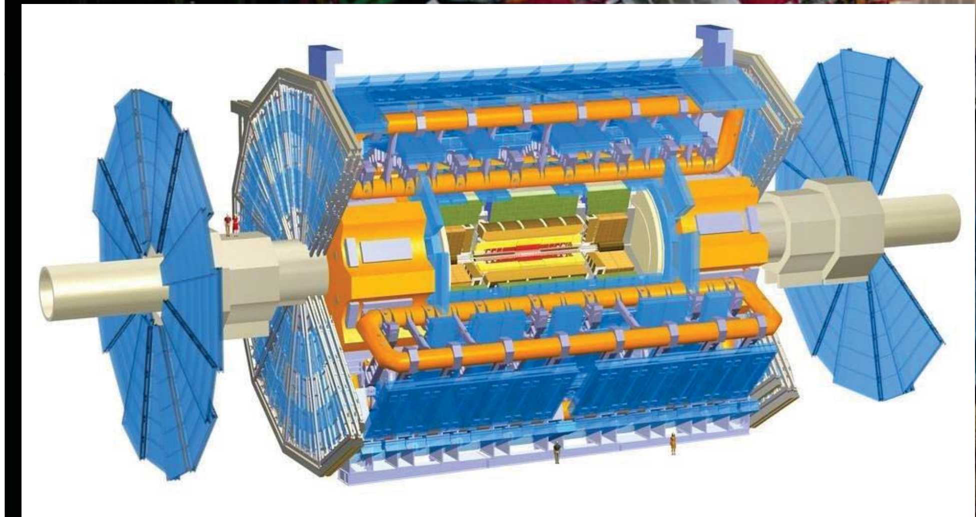


The Large Hadron Collider (LHC)

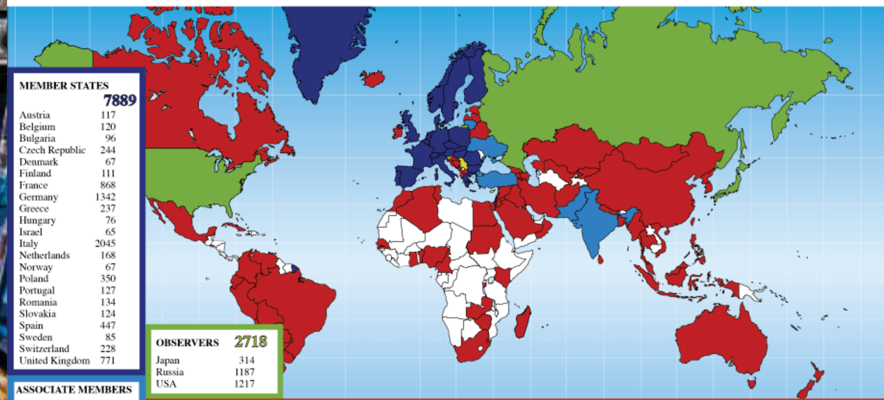


A new frontier in Energy & **Data volumes**:

LHC experiments generate 50 PB/year in Run 2



Distribution of All CERN Users by Nationality on 24 January 2018



MEMBER STATES 7889

Austria	117
Belgium	120
Bulgaria	96
Czech Republic	244
Denmark	67
Finland	111
France	868
Germany	1342
Greece	237
Hungary	76
Israel	65
Italy	2045
Netherlands	168
Norway	67
Poland	350
Portugal	127
Romania	134
Slovakia	124
Spain	447
Sweden	85
Switzerland	228
United Kingdom	771

ASSOCIATE MEMBERS 745

India	357
Lithuania	35
Pakistan	65
Turkey	173
Ukraine	115

ASSOCIATE MEMBERS IN THE PRE-STAGE TO MEMBERSHIP 118

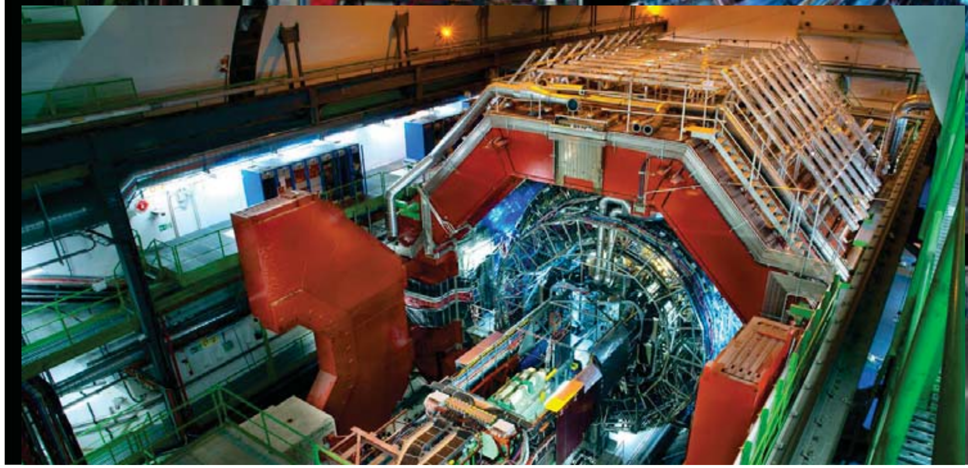
Cyprus	26
Serbia	57
Slovenia	35

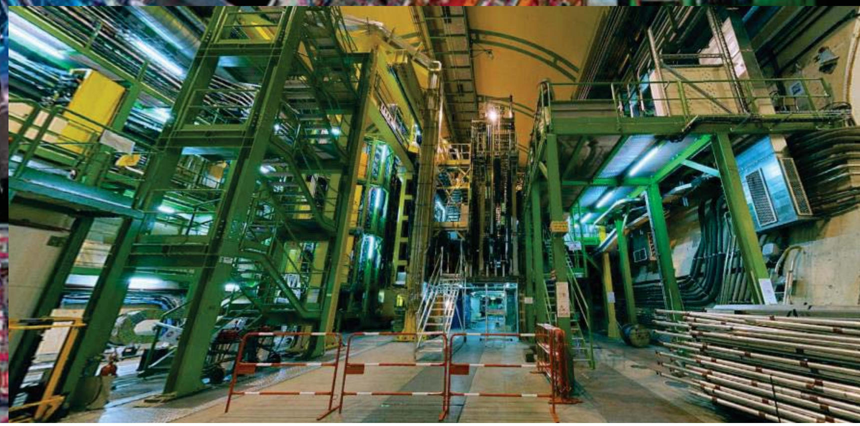
OBSERVERS 2718

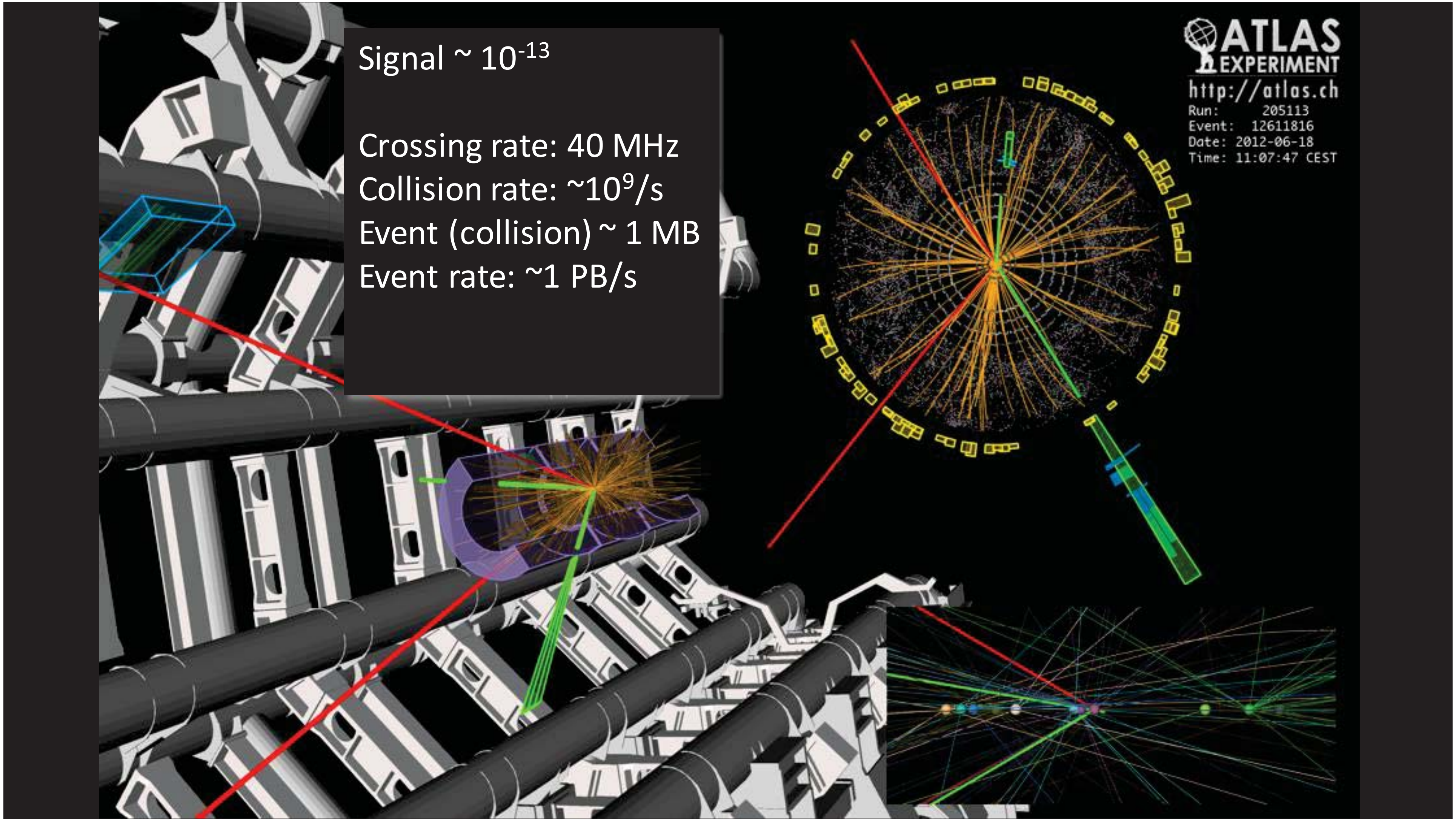
Japan	314
Russia	1187
USA	1217

OTHERS 1872

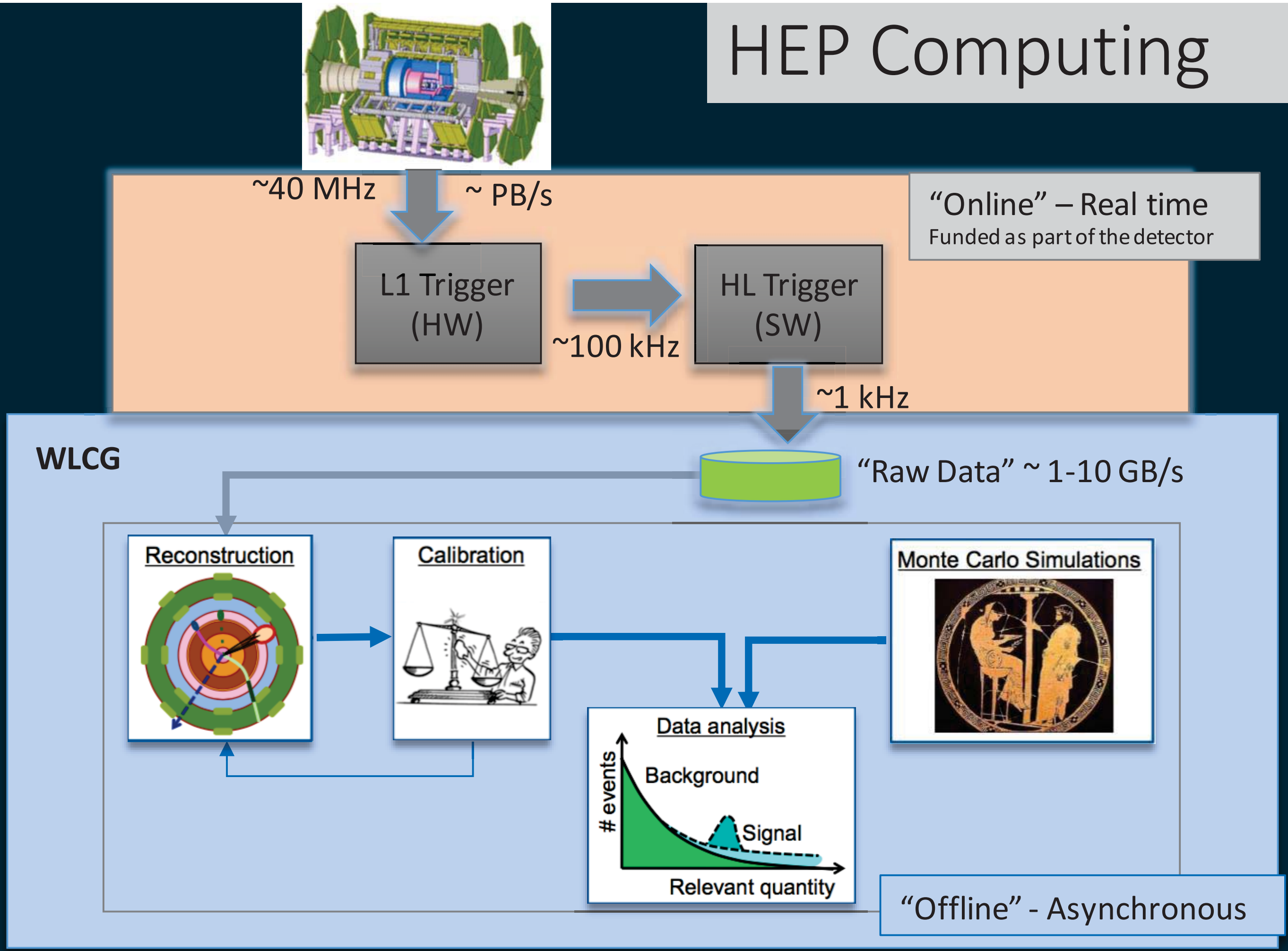
Afghanistan	1
Albania	3
Algeria	14
Argentina	27
Armenia	19
Australia	31
Azerbaijan	10
Bangladesh	11
Belarus	48
Benin	1
Bolivia	4
Bosnia & Herzegovina	2
Brazil	135
Burundi	1
Cameroon	1
Canada	161
Chile	20
China	510
Colombia	45
Croatia	41
Cuba	12
Ecuador	6
Egypt	31
El Salvador	1
Estonia	15
Georgia	46
Ghana	1
Hong Kong	1
Iceland	3
Indonesia	11
Iran	51
Iraq	1
Ireland	16
Jordan	1
Kazakhstan	5
Kenya	3
Korea Rep.	185
Kyrgyzstan	1
Latvia	2
Lebanon	23
Luxembourg	2
Madagascar	4
Malaysia	15
Malta	9
Mauritius	1
Mexico	82
Mongolia	2
Montenegro	11
Morocco	20
Myanmar	1
Nepal	10
New Zealand	5
Nigeria	3
North Korea	1
Oman	3
Palestine (O.T.).	7
Paraguay	2
Peru	7
Philippines	3
Saint Kitts and Nevis	1
Saudi Arabia	2
Senegal	1
Singapore	4
South Africa	56
Sri Lanka	6
Sudan	1
Swaziland	1
Syria	1
Taiwan	51
Thailand	22
T.F.Y.R.O.M.	2
Tunisia	5
Uruguay	1
Uzbekistan	4
Venezuela	10
Viet Nam	13
Zambia	1
Zimbabwe	2

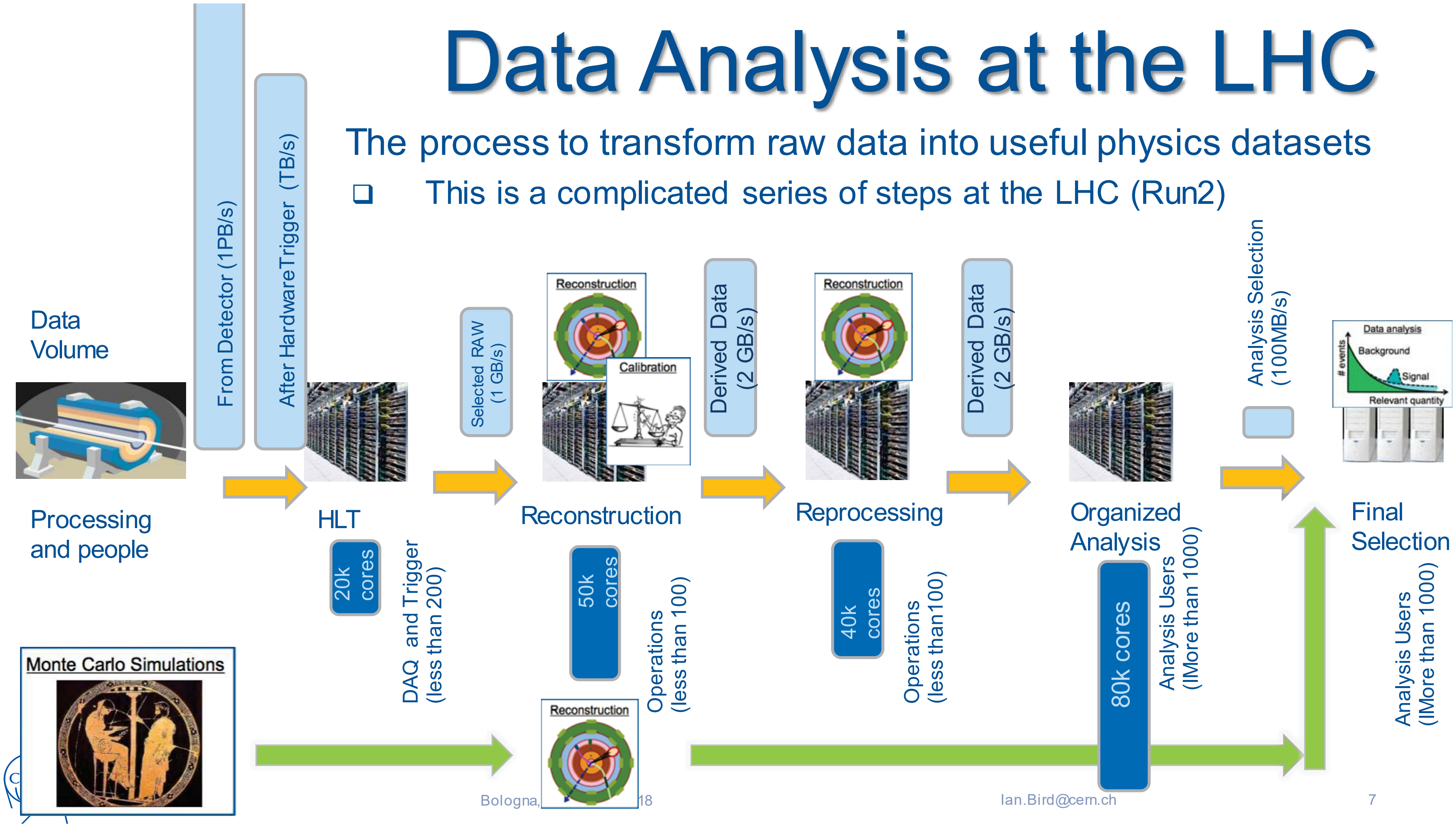





Signal ~ 10^{-13}
Crossing rate: 40 MHz
Collision rate: ~10^9/s
Event (collision) ~ 1 MB
Event rate: ~1 PB/s
ATLAS EXPERIMENT
http://atlas.ch
Run: 205113
Event: 12611816
Date: 2012-06-18
Time: 11:07:47 CEST

HEP Computing

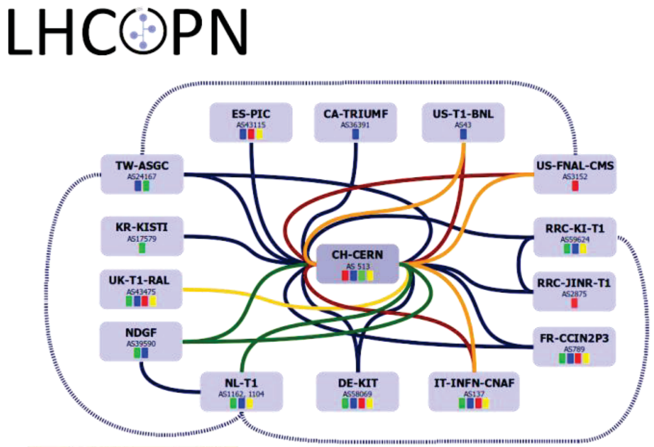


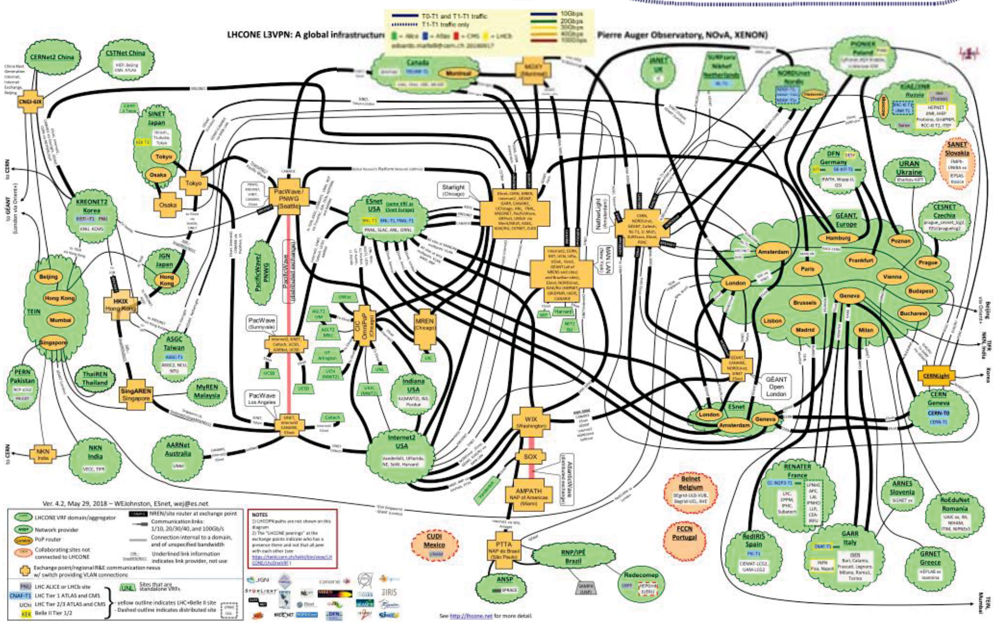


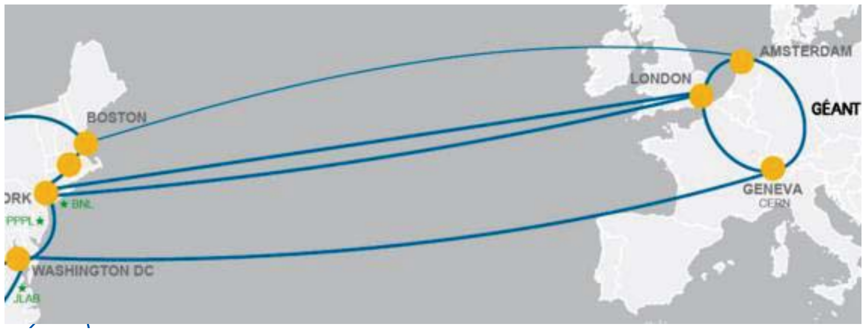
Data Analysis at the LHC
The process to transform raw data into useful physics datasets
This is a complicated series of steps at the LHC (Run2)
Data Volume
From Detector (1PB/s)
After Hardware Trigger (TB/s)
Selected RAW (1 GB/s)
Derived Data (2 GB/s)
Derived Data (2 GB/s)
Analysis Selection (100MB/s)
Reconstruction
Calibration
Reconstruction
Data analysis
Background
Signal
Relevant quantity
events
Processing and people
HLT
Reconstruction
Reprocessing
Organized Analysis
Final Selection
20k cores
DAQ and Trigger (less than 200)
50k cores
Operations (less than 100)
40k cores
Operations (less than 100)
80k cores
Analysis Users (More than 1000)
Analysis Users (More than 1000)
Monte Carlo Simulations
Reconstruction

Scale of computing needs

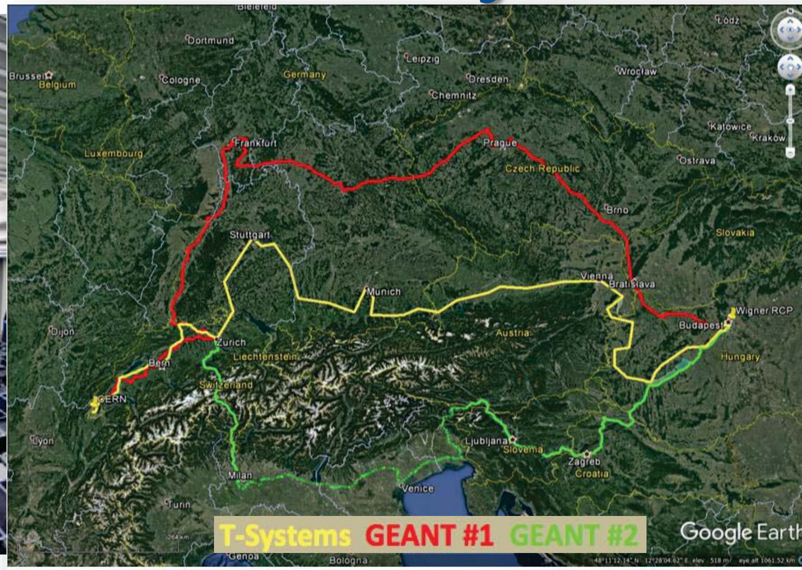
- CPU:
  - ~ 1 million cores fully occupied (“x86”)
- Storage
  - ~ 1 EB (~500 PB disk, >500 PB tape)
- Global networking
  - Some private 10-100 Gbps
  - LHCOne – overlay





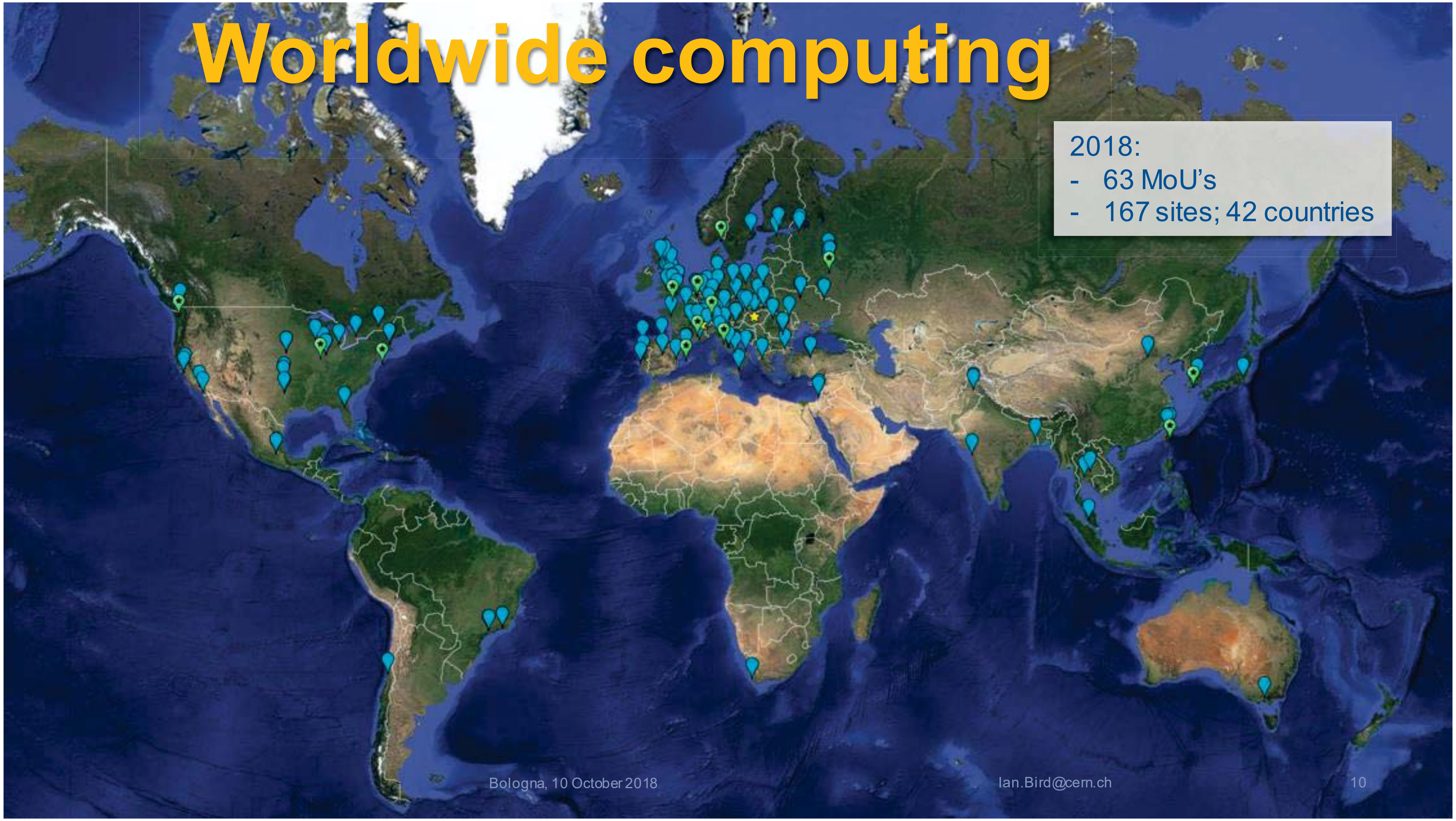


CERN Facilities today



COMPUTING		STORAGE	
Servers (Meyrin)	Cores (Meyrin)	Disks (Meyrin)	Tape Drives
11.5 K	174.3 K	61.9 K	104
Servers (Wigner)	Cores (Wigner)	Disks (Wigner)	Tape Cartridges
3.5 K	56.0 K	29.7 K	32.2 K

~180 PB usable disk
~250 PB on tape


Worldwide computing
2018:
- 63 MoU's
- 167 sites; 42 countries
Bologna, 10 October 2018
Ian.Bird@cern.ch
10

The Scale of the LHC Computing Problem

1 PB/s of data generated by the detectors
Up to **60 PB/year** of stored data

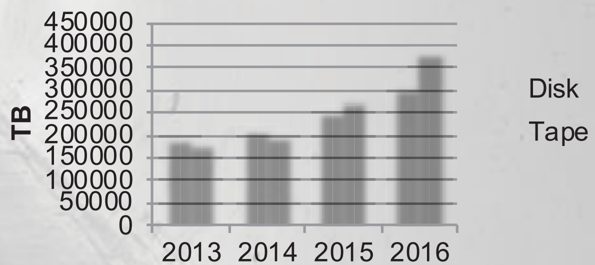
Large experiments have *managed* data sets of **>200 PB**

A distributed computing infrastructure of order of a **million cores** working 24/7
An average of 60M jobs/month

An continuous data transfer rate of 35-45 GB/s (**3 PB/day**) across the Worldwide LHC Grid (WLCG)

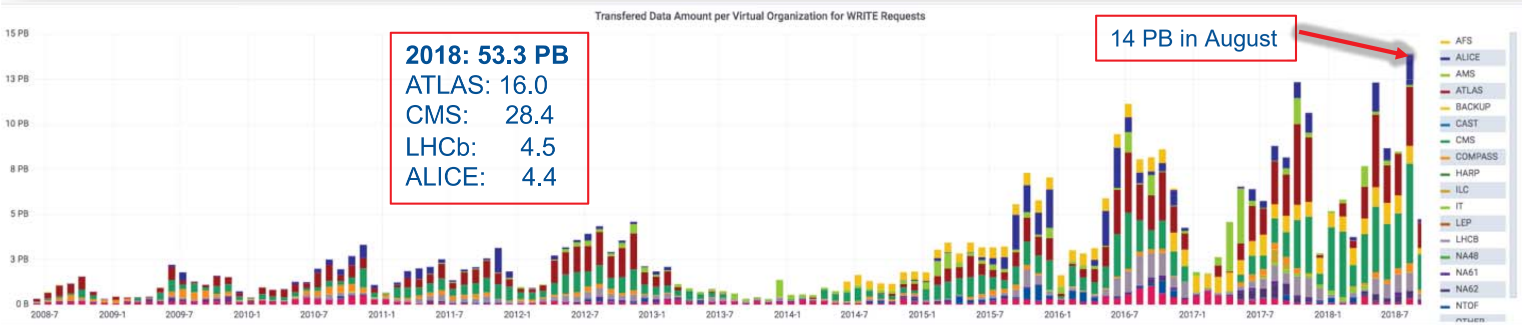
Would put us amongst the top Supercomputers if centrally placed: est. ~few x100 Pflops

More than 100 PB/month moved and accessed by 10k people

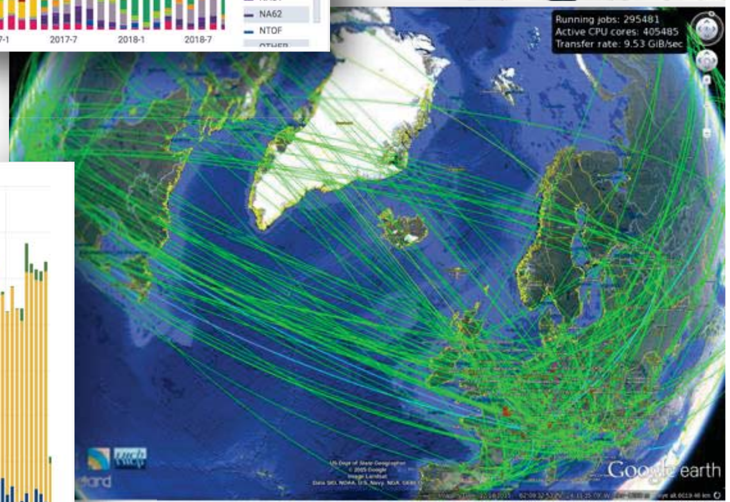


LHC Disk and Tape Storage

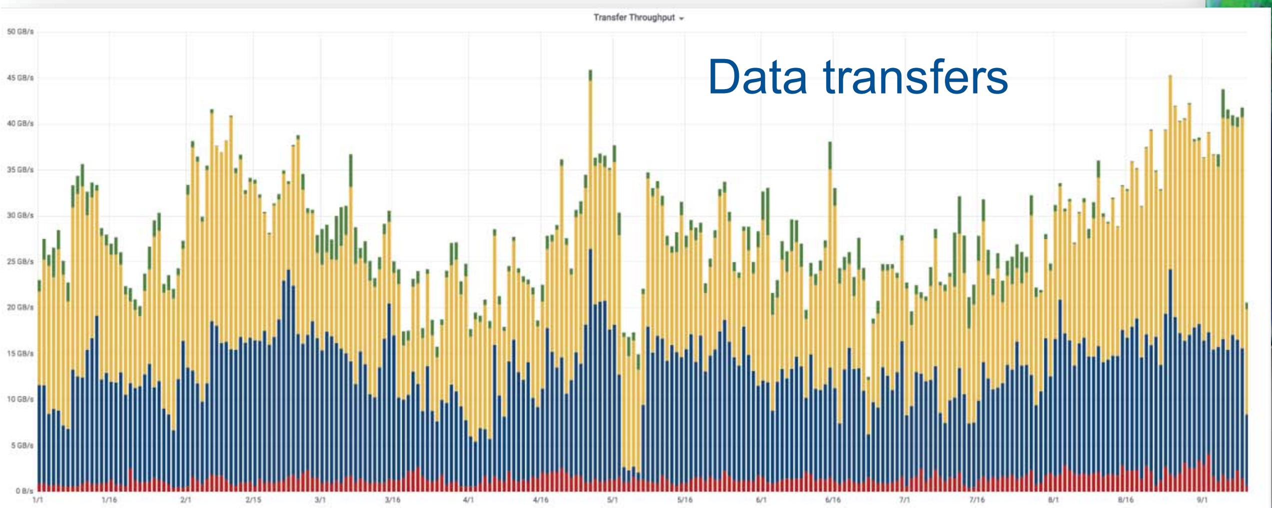
Data



Transfered Data Amount per Virtual Organization for WRITE Requests
2018: 53.3 PB
ATLAS: 16.0
CMS: 28.4
LHCb: 4.5
ALICE: 4.4
14 PB in August
AFS
ALICE
AMS
ATLAS
BACKUP
CAST
CMS
COMPASS
HARP
ILC
IT
LEP
LHCB
NA48
NA61
NA62
NTOF

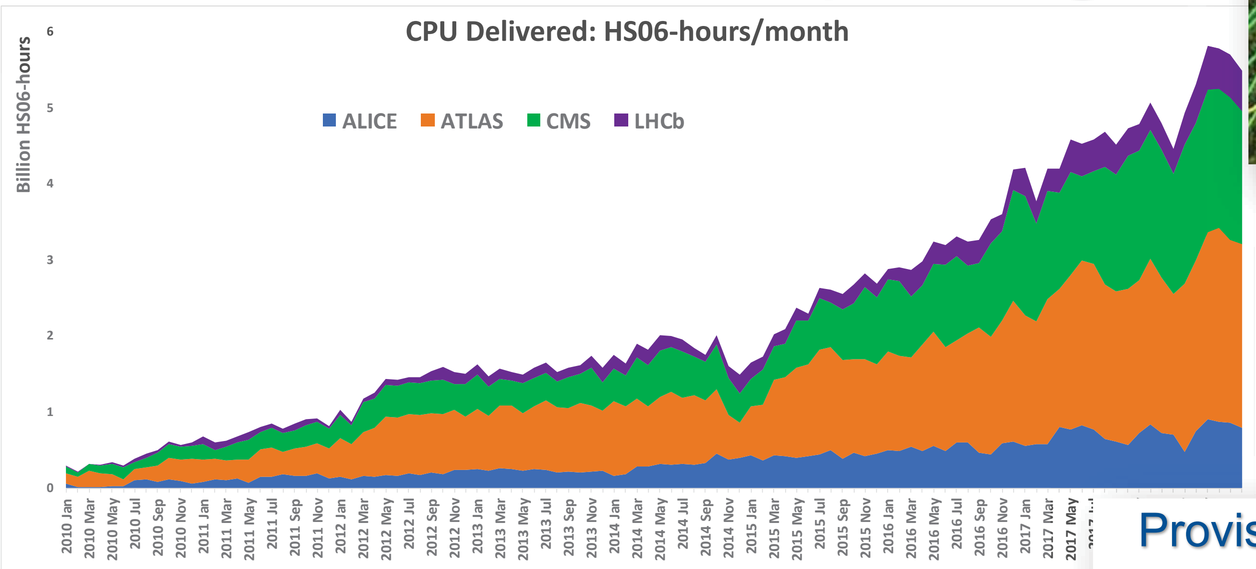


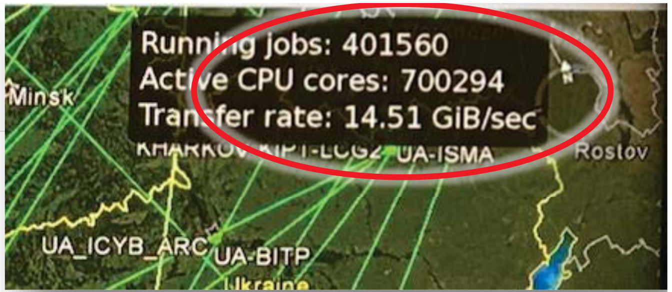
Running jobs: 295481
Active CPU cores: 405485
Transfer rate: 9.53 GiB/sec
Google earth

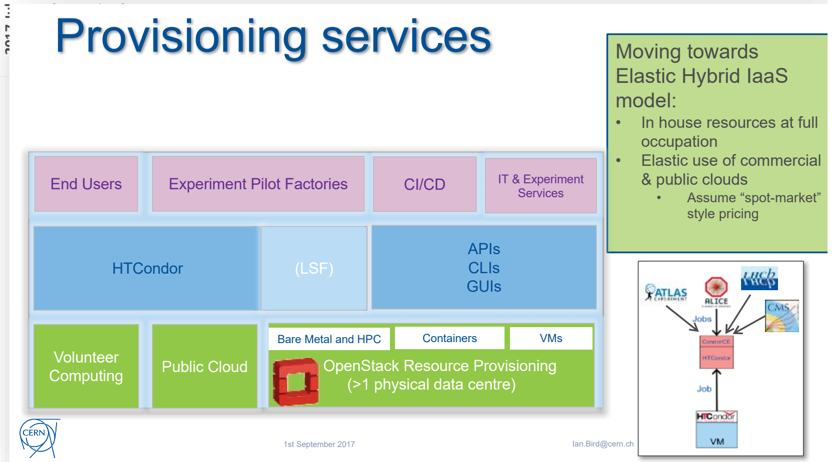


Transfer Throughput
Data transfers

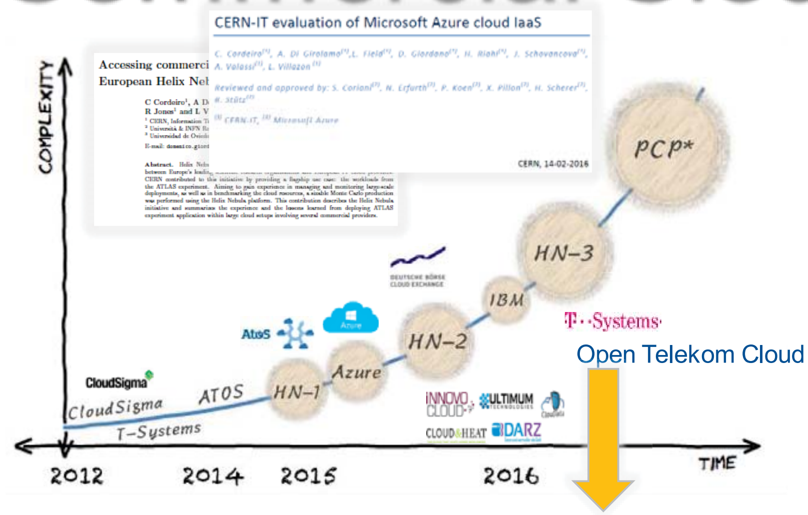
Worldwide computing

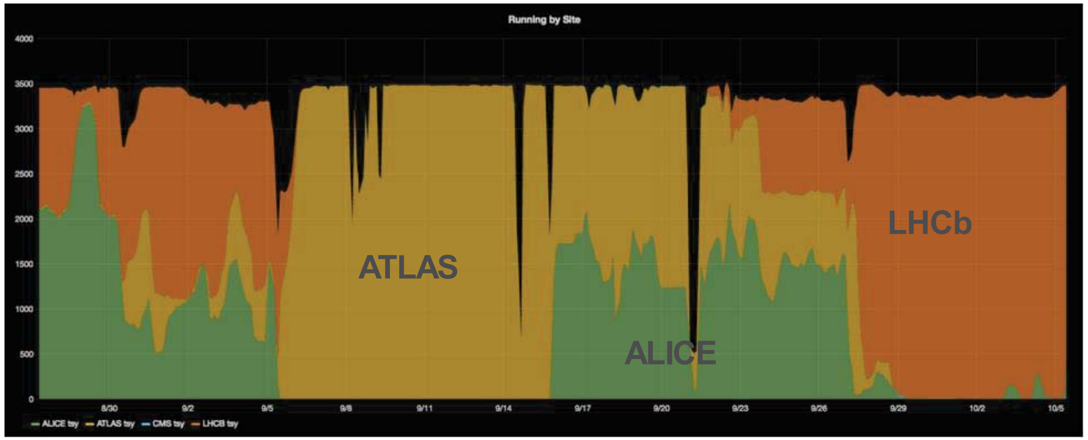


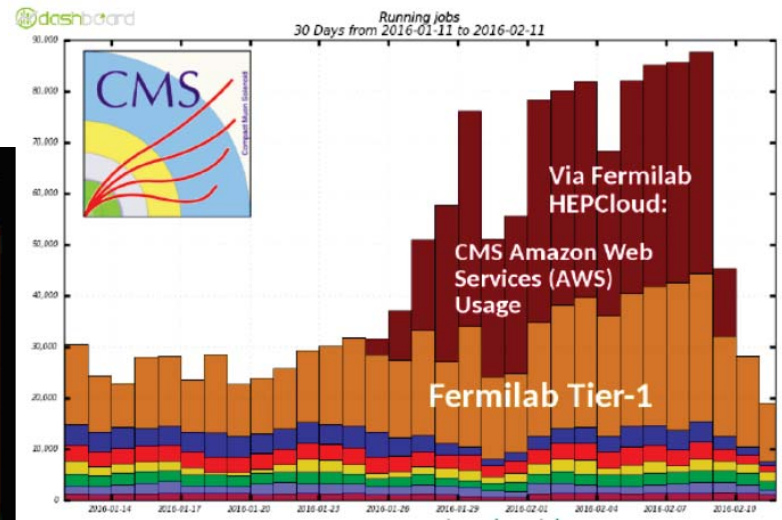




Commercial Clouds

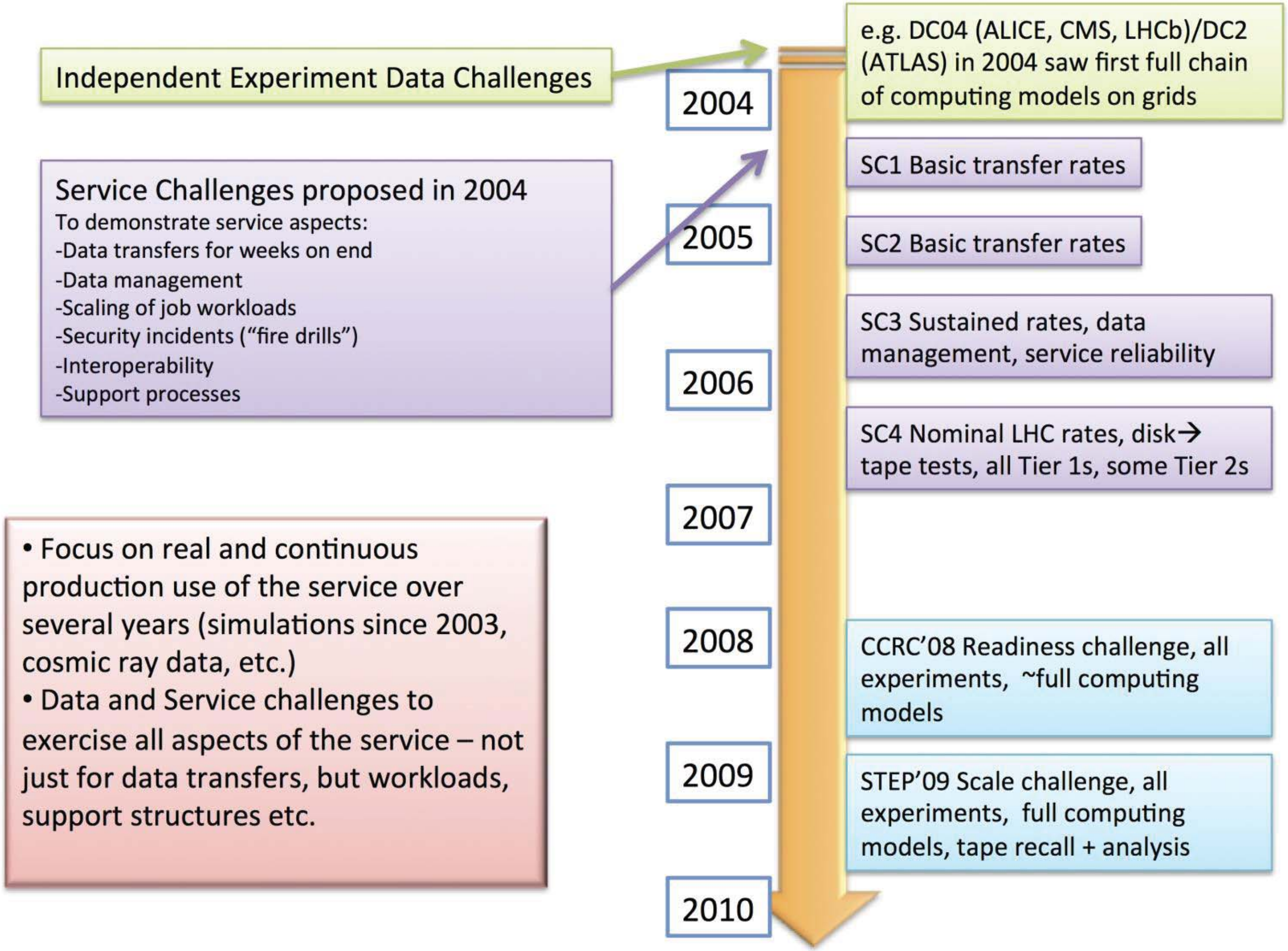






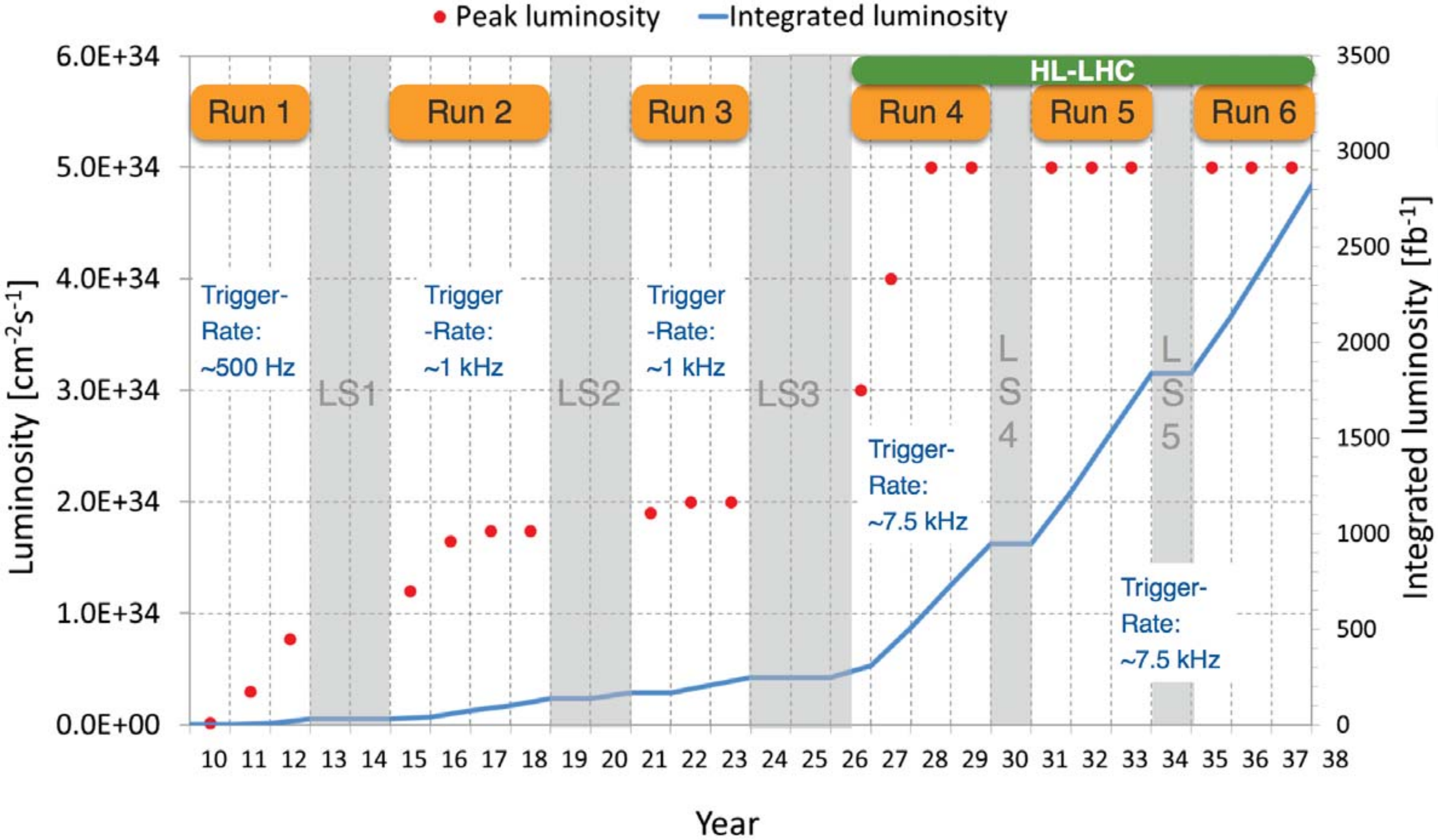
Data challenge programme

pre-LHC startup



CERN | WLCG

LHC Schedule



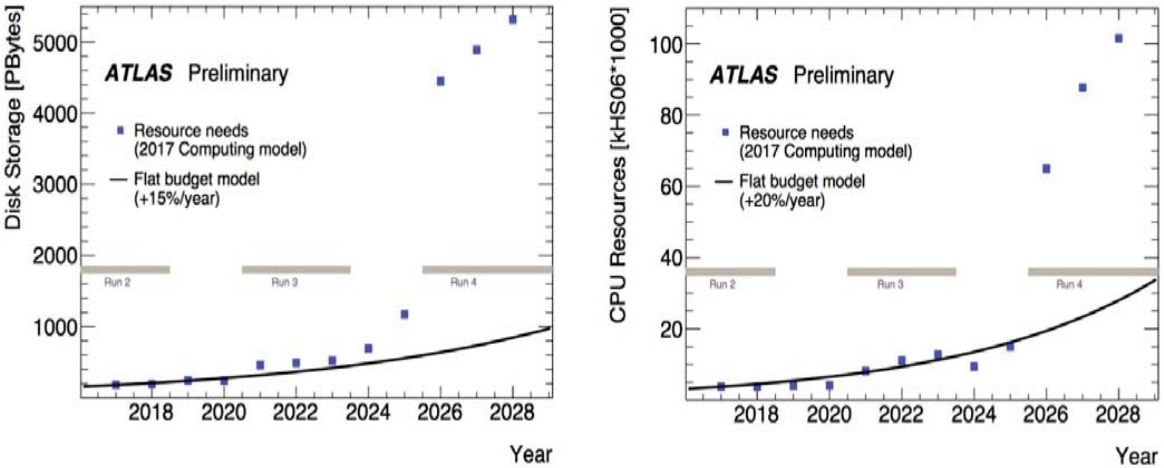
Run 3 Alice, LHCb upgrades

Run 4 ATLAS, CMS upgrades

The HL-LHC computing challenge

- HL-LHC needs for ATLAS and CMS are above the expected hardware technology evolution (15% to 20%/yr) and funding (flat)

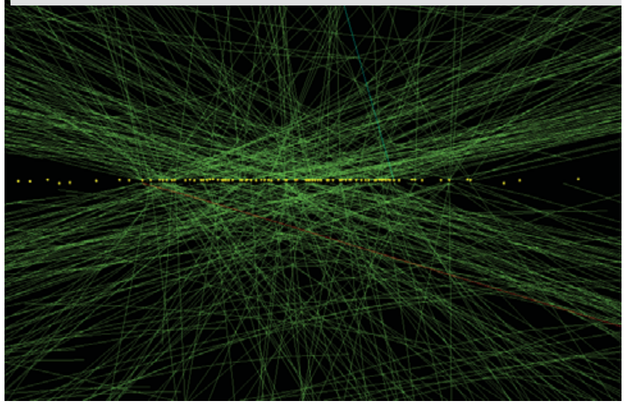
- The main challenge is storage, but computing requirements grow 20-50x



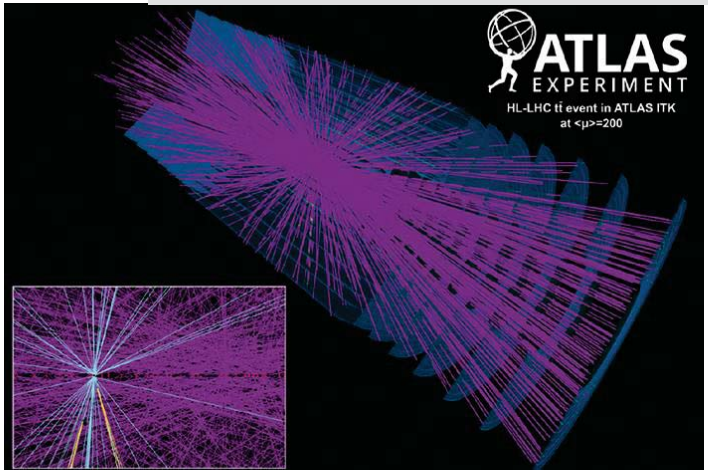
Events at HL-LHC

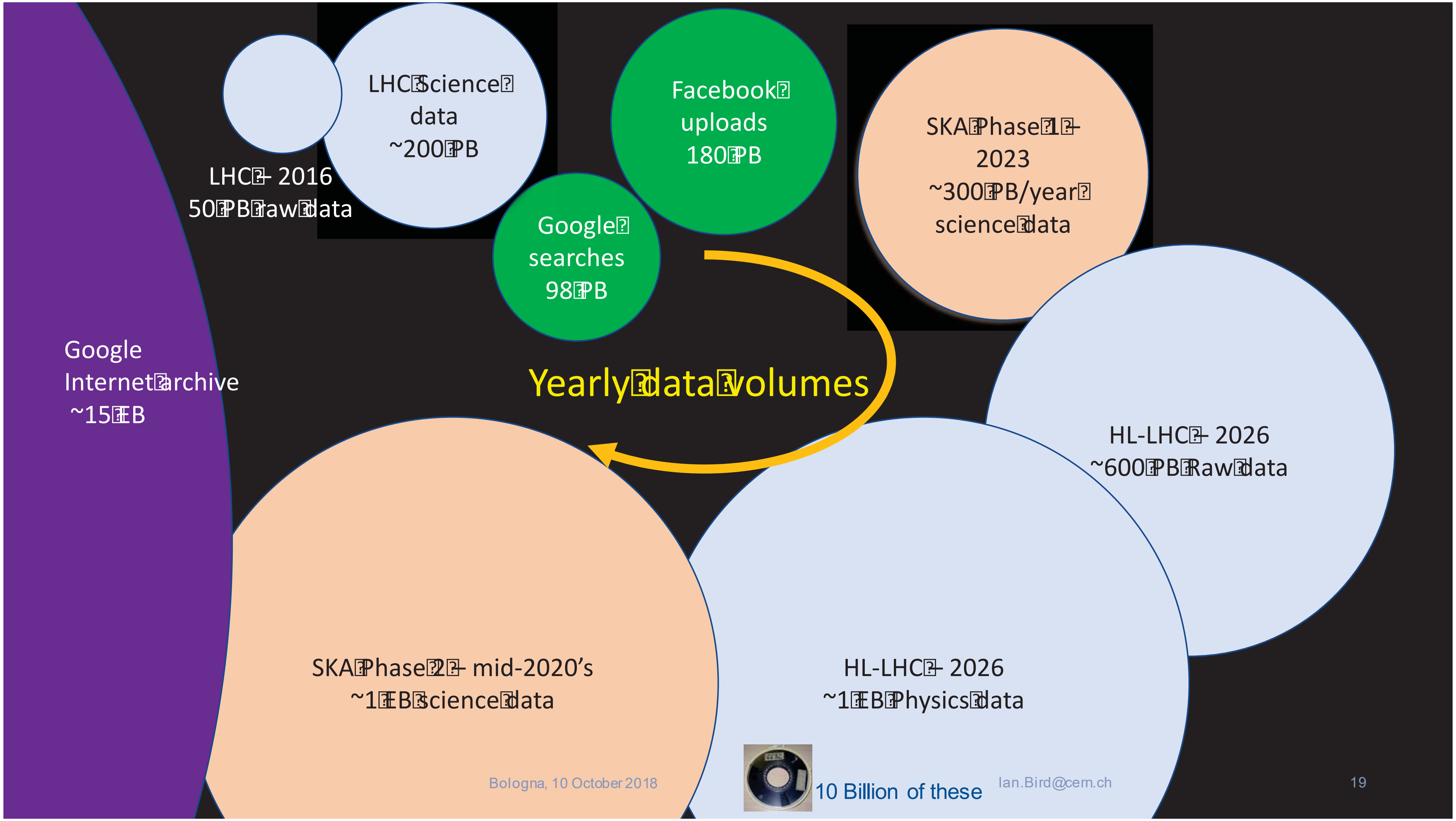
- Increased complexity due to much **higher pile-up** and higher trigger rates will bring several challenges to reconstruction algorithms

CMS: event from 2017 with 78 reconstructed vertices



ATLAS: simulation for HL-LHC with 200 vertices




Google Internet archive ~15 EB
LHC – 2016 50 PB raw data
LHC Science data ~200 PB
Facebook uploads 180 PB
Google searches 98 PB
SKA Phase 1 – 2023 ~300 PB/year science data
Yearly data volumes
HL-LHC – 2026 ~600 PB Raw data
SKA Phase 2 – mid-2020's ~1 EB science data
HL-LHC – 2026 ~1 EB Physics data
10 Billion of these

10-year challenges

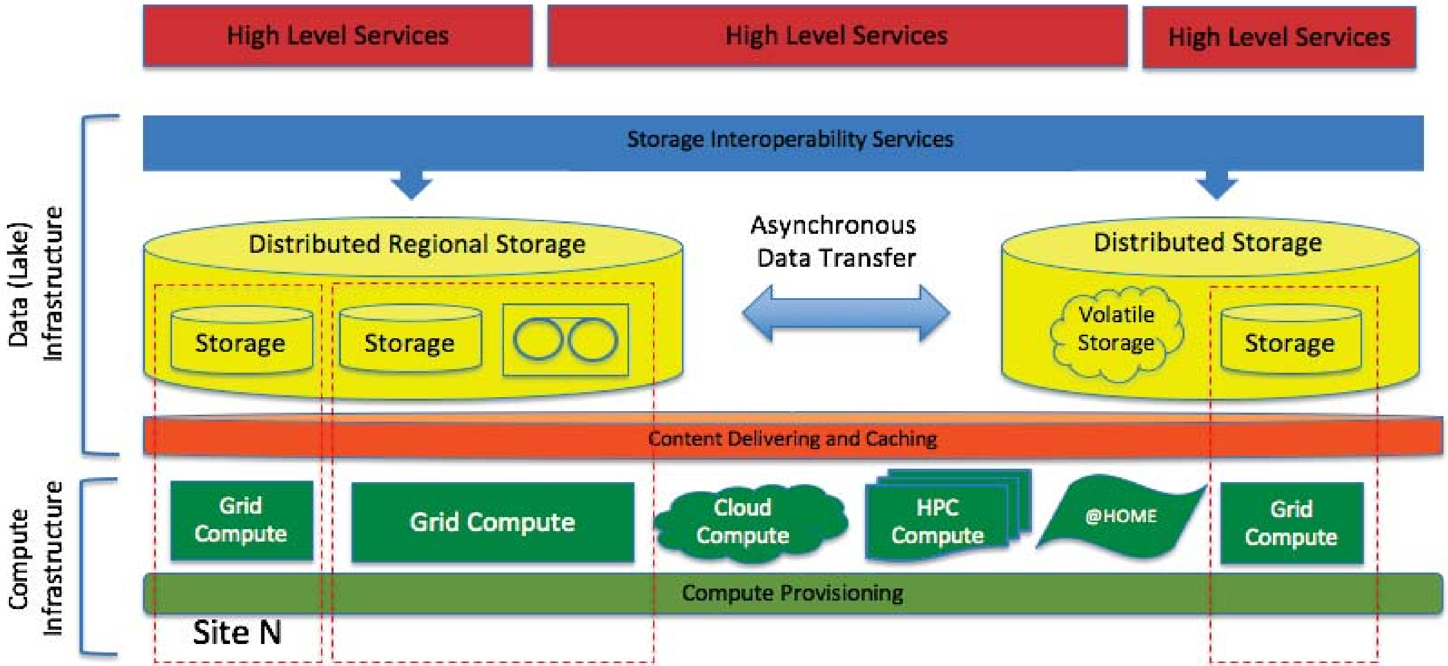
- HL-LHC will be a multi-Exabyte challenge
  - Storage and compute needs x10 above what naïve technology extrapolation will bring
  - Need to drive down costs: focus on performance, efficiency, operations, etc. ➔changes in computing and infrastructure models are necessary
- SKA will have similar data volumes on the same time-scale
- Opportunity for synergy – in particular in large scale facilities
  - SKA and LHC likely to be co-located in major facilities
- But there is experience:
  - ~15 years of grid development and successful operation for science
  - CERN has been operating a distributed DC for ~5 years
  - Large internet companies provide tools and experience that did not exist when we started WLCG
    - Tools for managing interconnected DCs, cloud provisioning, etc.
  - Starting to prototype federated structures for the future

Evolution of WLCG

- Community White Paper
  - 1 year – bottom up review of LHC computing topics
  - 13 working groups on all aspects
  - Outlines how HEP computing could evolve to address computing challenges
  - https://arxiv.org/abs/1712.06982
- WLCG Strategy Document
  - Prioritisation of topics in the CWP from the point of view of the HL-LHC challenges
  - Set out a number of R&D projects for the next 5 years
    - Running global system should evolve towards HL-LHC
  - http://cern.ch/go/Tg79

- Main R&D topics
  - Software performance, re-engineering, algorithmic improvement
    - New techniques, e.g. ML/DL
  - Evolution of data management, access, organization
    - Data lakes, transfer tools, protocols, acces mechanisms, caching, etc.
  - Integration of heterogenous compute:
    - Architectures, HPC, cloud, etc.
  - Cost and technology evolution – optimizing hardware cost
    - Reduction of data volumes
  - Managing operational costs

Conceptual view of “data lake”



Idea is to localize bulk data in a cloud service (Tier 1’s ➔ data lake): minimize replication, assure availability

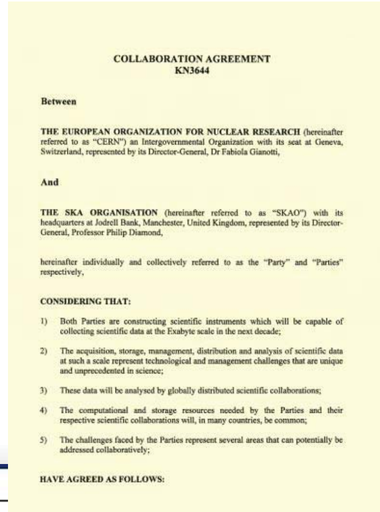
Serve data to remote (or local) compute – grid, cloud, HPC, ???

Simple caching is all that is needed at compute site

Works at national, regional, global scales

Collaboration CERN – SKA

- Recognition on both sides of potential synergies and requirements
  - Various ad-hoc interactions between communities
  - Reviews and panels etc.
  - Recently held a CERN-SKA “Big data” workshop in the UK Alan Turing Inst.
- In July 2017 CERN and SKAO DG’s signed a collaboration agreement on computing, data management, etc.
  - Recognizing that both HL-LHC and SKA will be Exabyte-scale scientific experiments on a 10-year timescale



COLLABORATION AGREEMENT
KN3644

Between

THE EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (hereinafter referred to as “CERN”) an Intergovernmental Organization with its seat at Geneva, Switzerland, represented by its Director-General, Dr Fabiola Gianotti,

And

THE SKA ORGANISATION (hereinafter referred to as “SKAO”) with its headquarters at Jodrell Bank, Manchester, United Kingdom, represented by its Director-General, Professor Philip Diamond,

hereinafter individually and collectively referred to as the “Party” and “Parties” respectively,

CONSIDERING THAT:

1) Both Parties are constructing scientific instruments which will be capable of collecting scientific data at the Exabyte scale in the next decade;

2) The acquisition, storage, management, distribution and analysis of scientific data at such a scale represent technological and management challenges that are unique and unprecedented in science;

3) These data will be analysed by globally distributed scientific collaborations;

4) The computational and storage resources needed by the Parties and their respective scientific collaborations will, in many countries, be common;

5) The challenges faced by the Parties represent several areas that can potentially be addressed collaboratively;

HAVE AGREED AS FOLLOWS:

CERN COURIER

Aug 11, 2017

SKA and CERN co-operate on extreme computing

Big-data co-operation agreement

ESFRI Science Projects

HL-LHC	SKA
FAIR	CTA
KM3Net	JIVE-ERIC
ELT	EST
EURO-VO	EGO-VIRGO
(LSST)	(CERN,ESO)

ESCAPE

European Science Cluster of Astronomy & Particle physics
ESFRI research infrastructures

Goals:

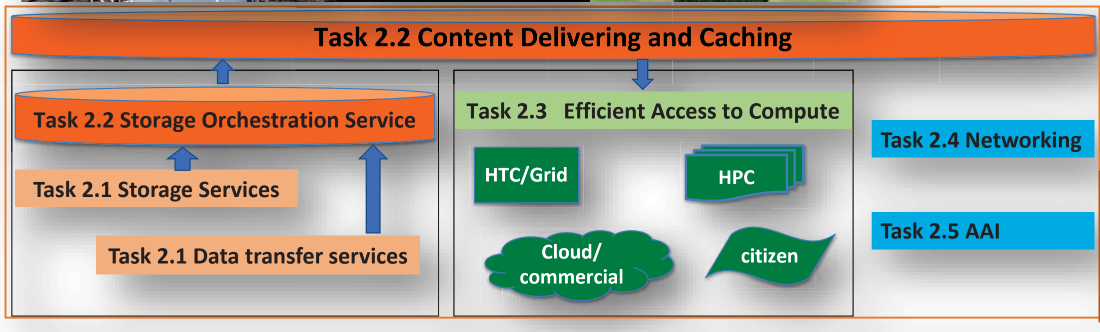
Prototype an infrastructure for the EOSC that is adapted to the Exabyte-scale needs of the large ESFRI science projects.

Ensure that the science communities drive the development of the EOSC.

Has to address *FAIR* data management, long term preservation, open access, open science, and contribute to the EOSC catalogue of services.

Work Packages

WP2 – Data Infrastructure for Open Science
WP3 – Open-source scientific Software and Service Repository
WP4 – Connecting ESFRI projects to EOSC through VO framework
WP5 – ESFRI Science Analysis Platform



Data centres (funded in WP2)
CERN, INFN, DESY, GSI, Nikhef, SURFSara, RUG, CCIN2P3, PIC, LAPP, INAF

HPC?

- **HPCs are here** in HEP computing, here to stay and grow
  - They require **dedicated investment** of effort
  - We require **stable allocations**, not just backfill, to make the investments pay; resource acquisition model is important
- They bring **accelerators** like GPUs with them, which we can't leave idle
- Particularly crucial for ATLAS and CMS: on an HL-LHC timescale, major funding agencies are mandating a **very high profile for HPCs**
  - Beginning with a mandate to use the first exascale machine in the US in 2021
- We've done the preparatory work for using accelerators in simu/reco, building multithreaded frameworks, but we seem **far from applications that are exascale ready**
- Are **machine learning applications** -- or at least their training component -- the most achievable path to apps for the first exascale machine in 2021?
  - Can we sketch out now more ambitious objectives for Run-4?
- Many questions to answer: we must **boost our development efforts and enlist CS experts** to help answer them

BROOKHAVEN T. Wenaus September 2018 24

- Yes, BUT:
  - HPC not designed for our applications, so not what we would choose to use
  - Each machine is a one-off, no common environment (software, usage)
  - Need federated identity support, and reasonable security environment
  - Need external connectivity …
  - Need real support for Exabyte-scale data processing (getting data to each core)
  - Only certain applications can (will ever?) make use of accelerators
  - Accelerated hardware available but it is hard to adopt
  - Need for serious modifications in the allocation model to get available and sustained resources