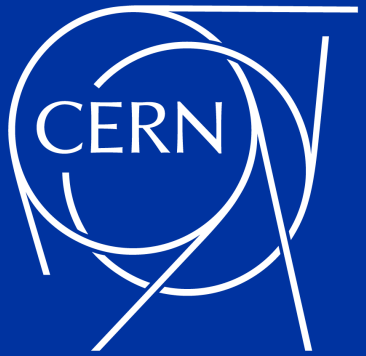


HL-LHC network challenges

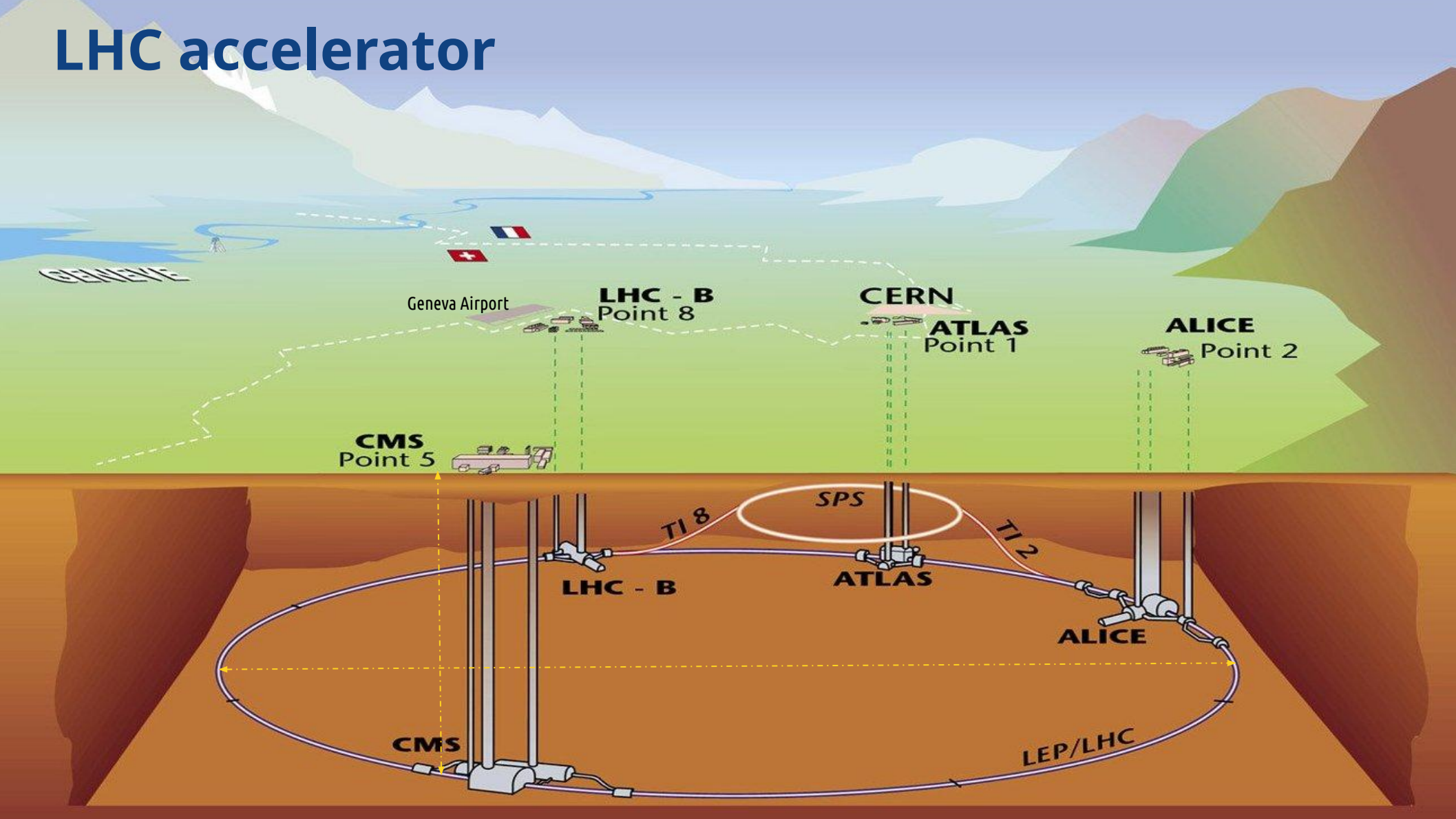
TNC25 Brighton - UK
June 2025
Eric Grancher, Edoardo Martelli - CERN

Agenda

- CERN and HL-LHC
- WLCG and Data Challenges
- Network R&D

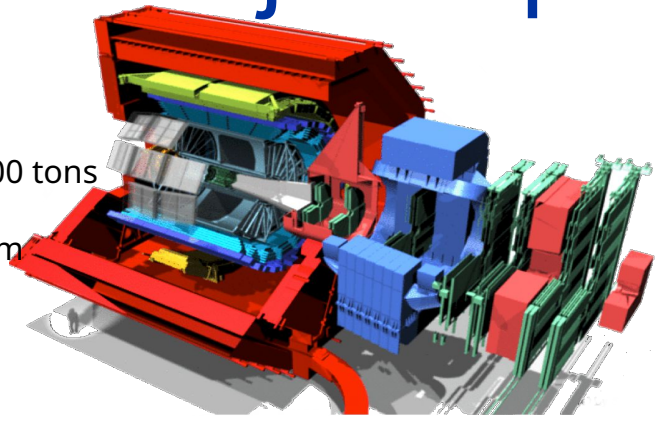
CERN and HL-LHC

LHC accelerator

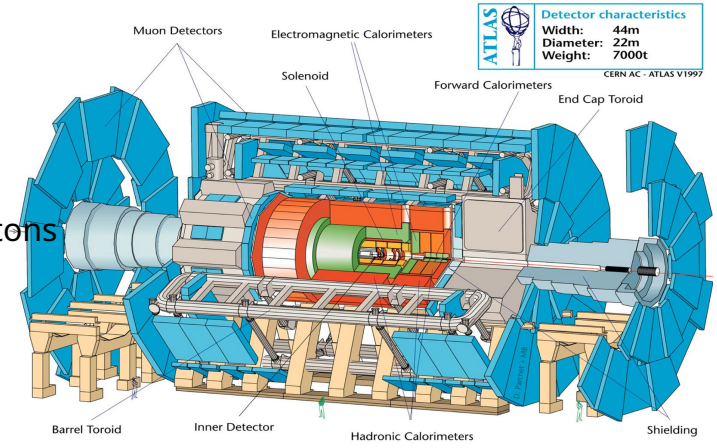


LHC major experiments

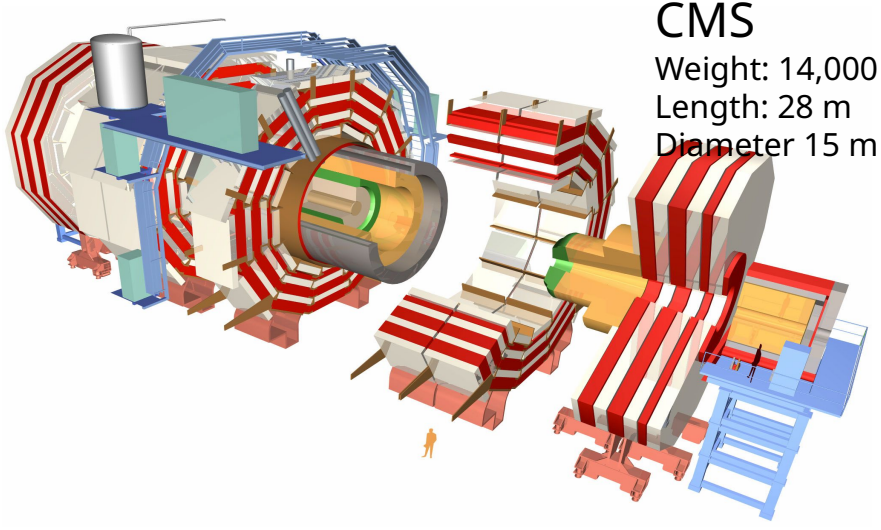
ALICE
Weight: 10,000 tons
Length: 26 m
Diameter 16 m



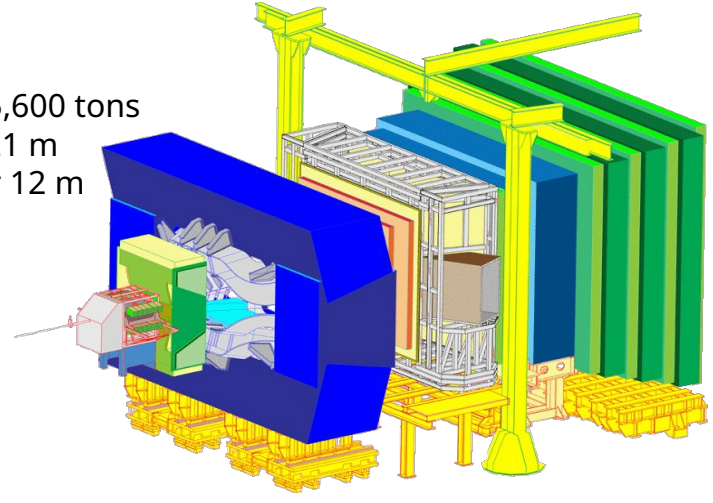
ATLAS
Weight: 7,000 tons
Length: 44 m
Diameter 22 m



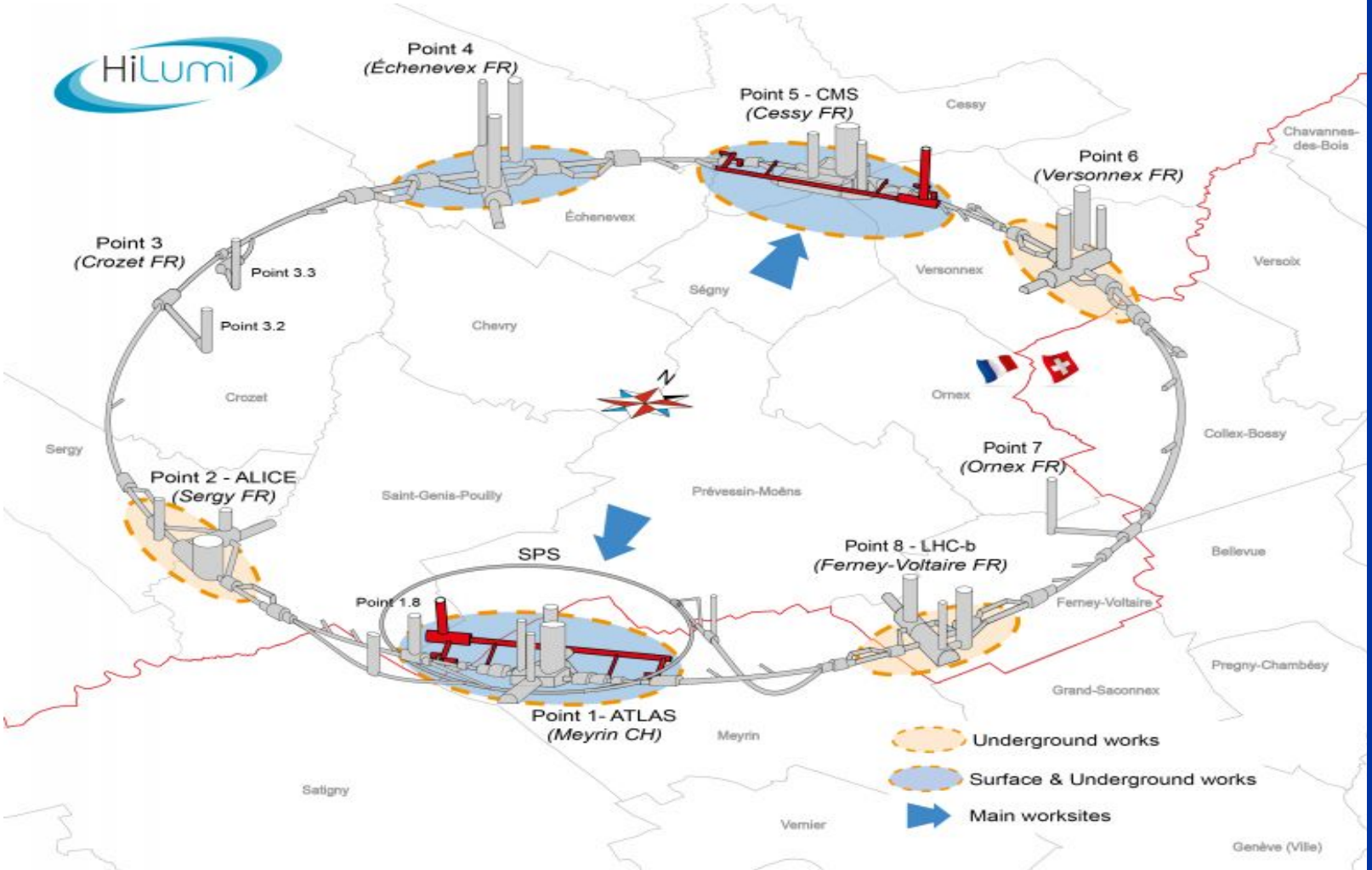
CMS
Weight: 14,000 tons
Length: 28 m
Diameter 15 m



LHCb
Weight: 5,600 tons
Length: 21 m
Diameter 12 m



Next: the HL-LHC project

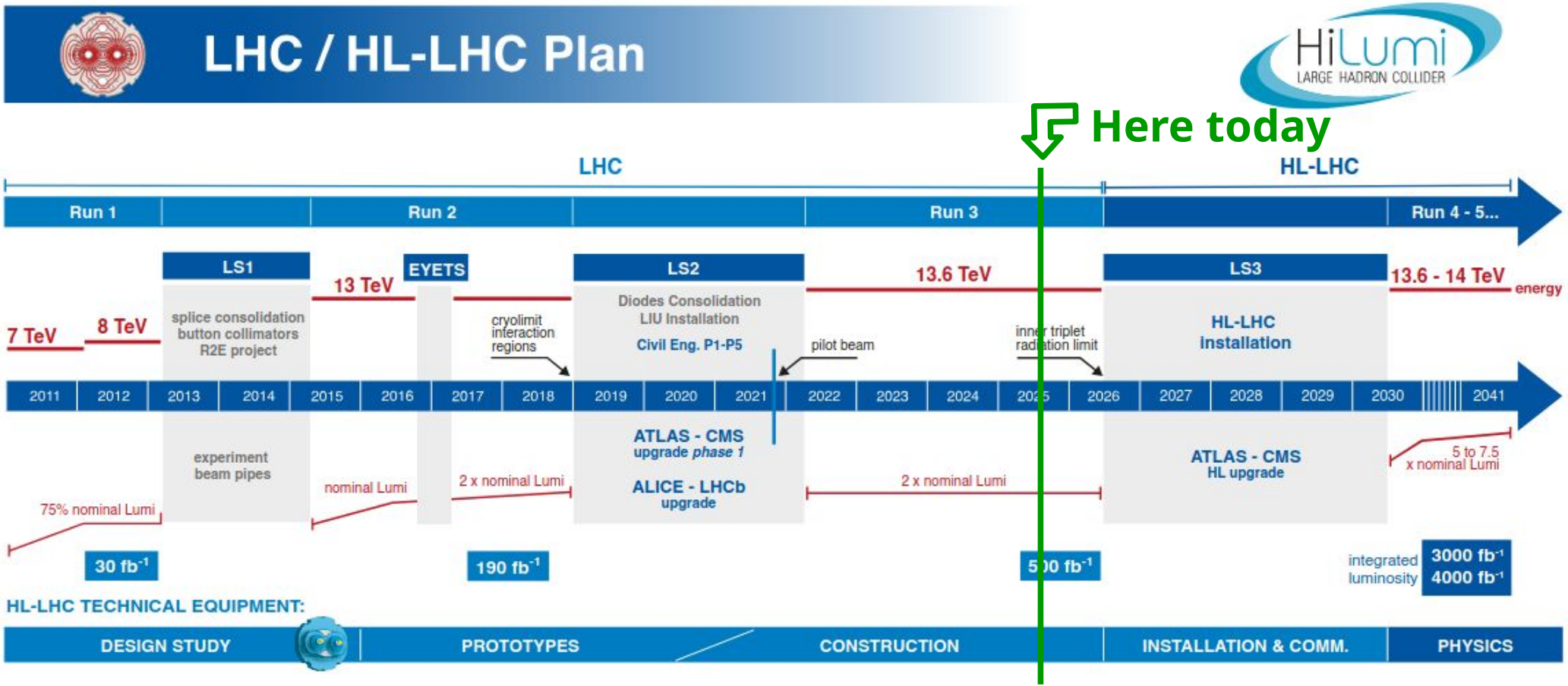


The **High-Luminosity Large Hadron Collider** (HL-LHC) is an **upgraded version of the LHC**

It will operate at a higher luminosity, i.e. it will produce more collisions and data

The HL-LHC will enter service in 2030, **increasing the volume of data** produced by the experiments **by a factor of 10**

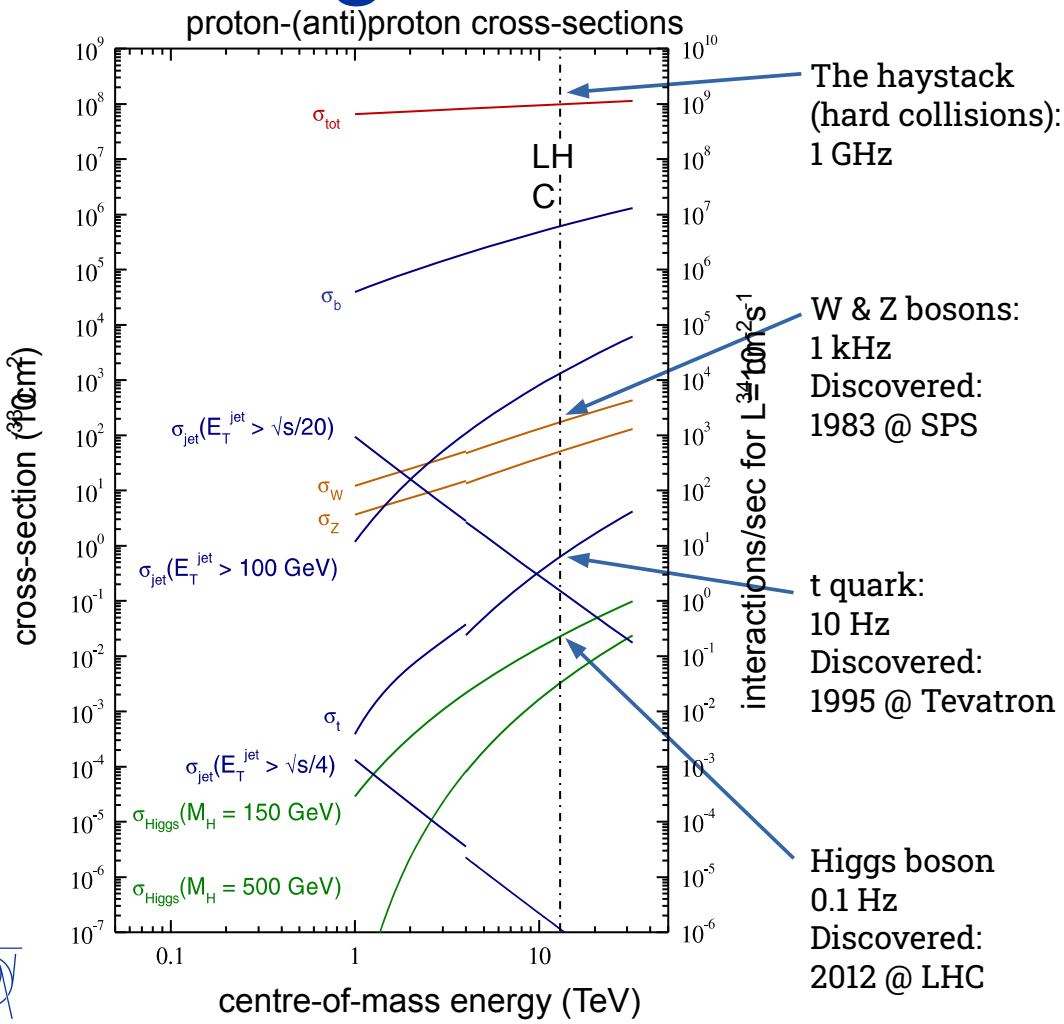
LHC / HL-LHC Plan



https://hilumilhc.web.cern.ch/content/hl-lhc-project

HL-LHC construction works

Finding the needle in the haystack



The haystack (hard collisions): 1 GHz

W & Z bosons: 1 kHz
Discovered: 1983 @ SPS

t quark: 10 Hz
Discovered: 1995 @ Tevatron

Higgs boson 0.1 Hz
Discovered: 2012 @ LHC

Contemporary particle physics: rare processes

The vast majority of the collisions are not interesting:

- they only result in particles and processes that were studied deeply decades ago

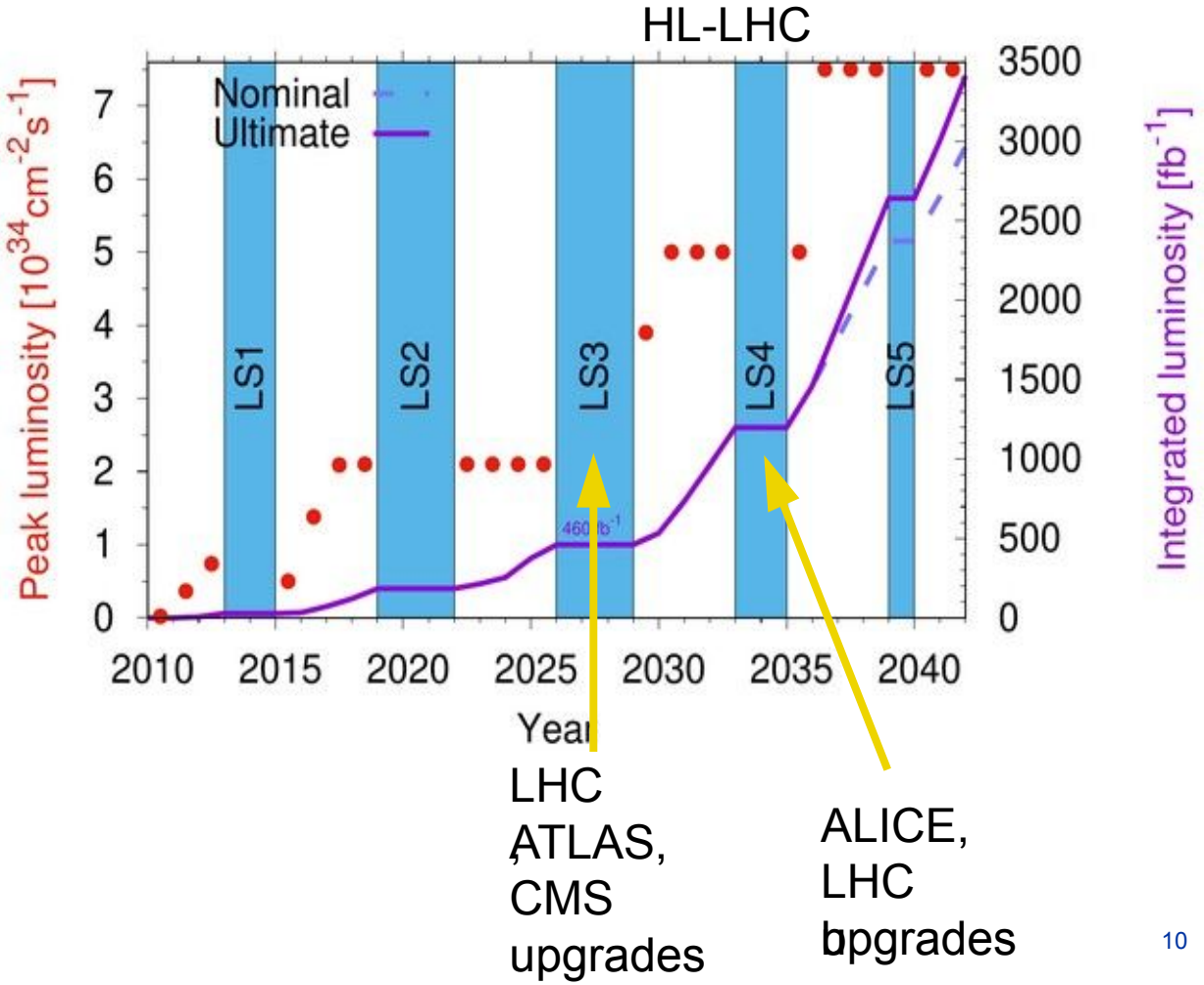
- interesting physics is >= 9 orders of magnitude rarer (i.e. >= one in a billion)

HL-LHC upgrade

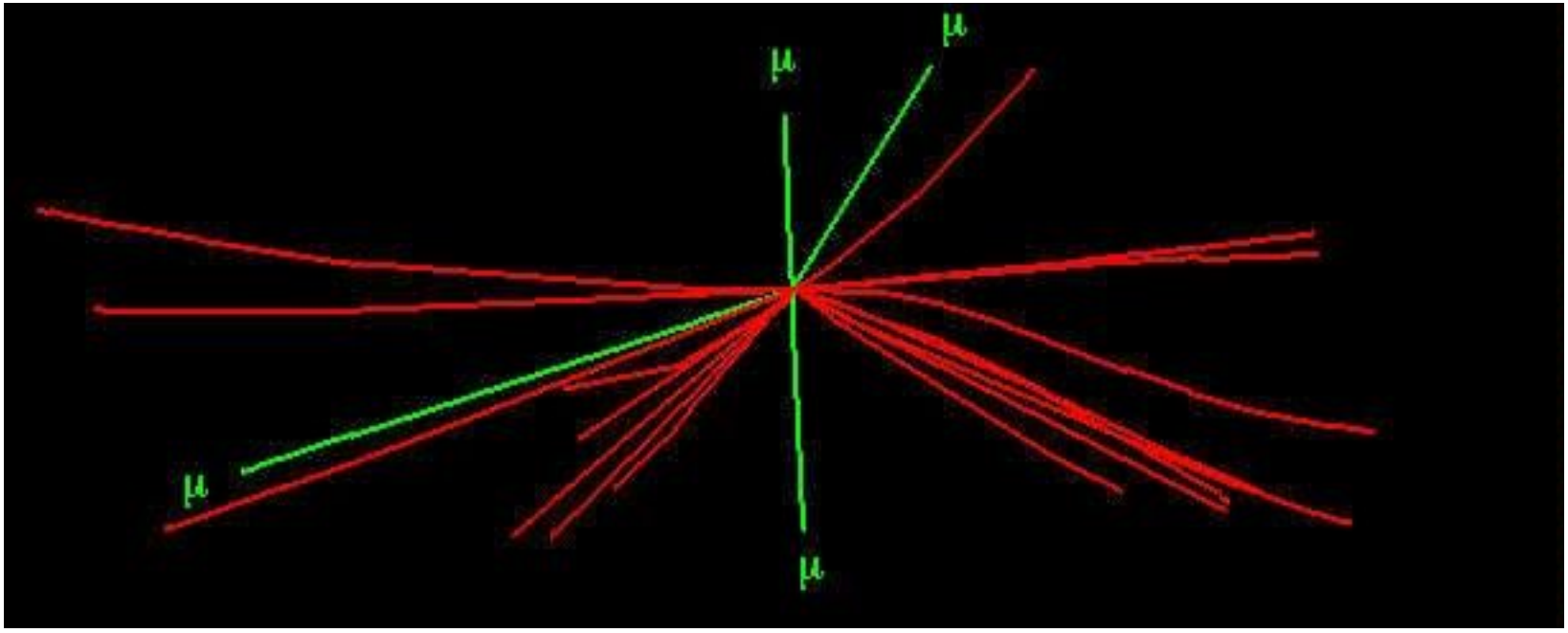
Luminosity: number of collisions

More collisions means more rare events

Goal: deliver 6-8 times more collisions in 11 years (2030-41) than LHC in 16 years (2010-26)



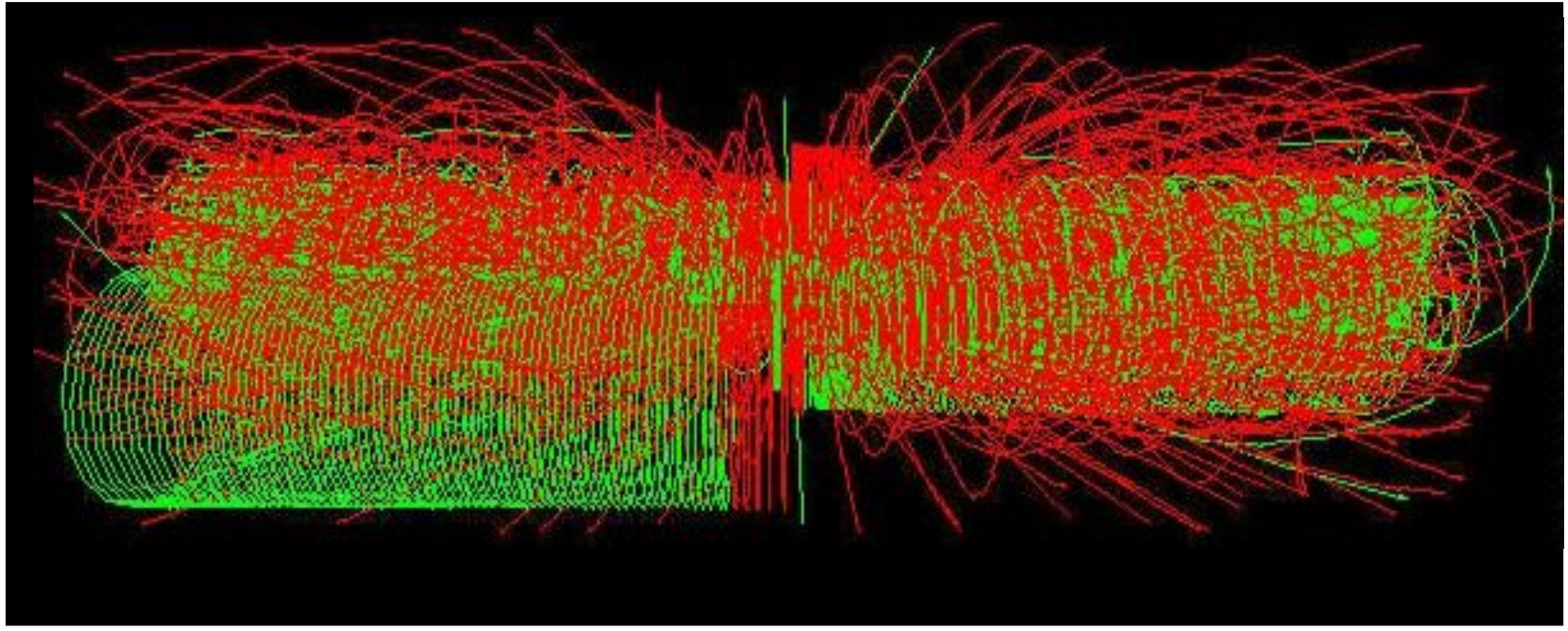
The needle



This is what we are looking for: a Higgs boson decaying in four easily identifiable muons

LHC produces a few of these per day, HL-LHC will increase 6-8x

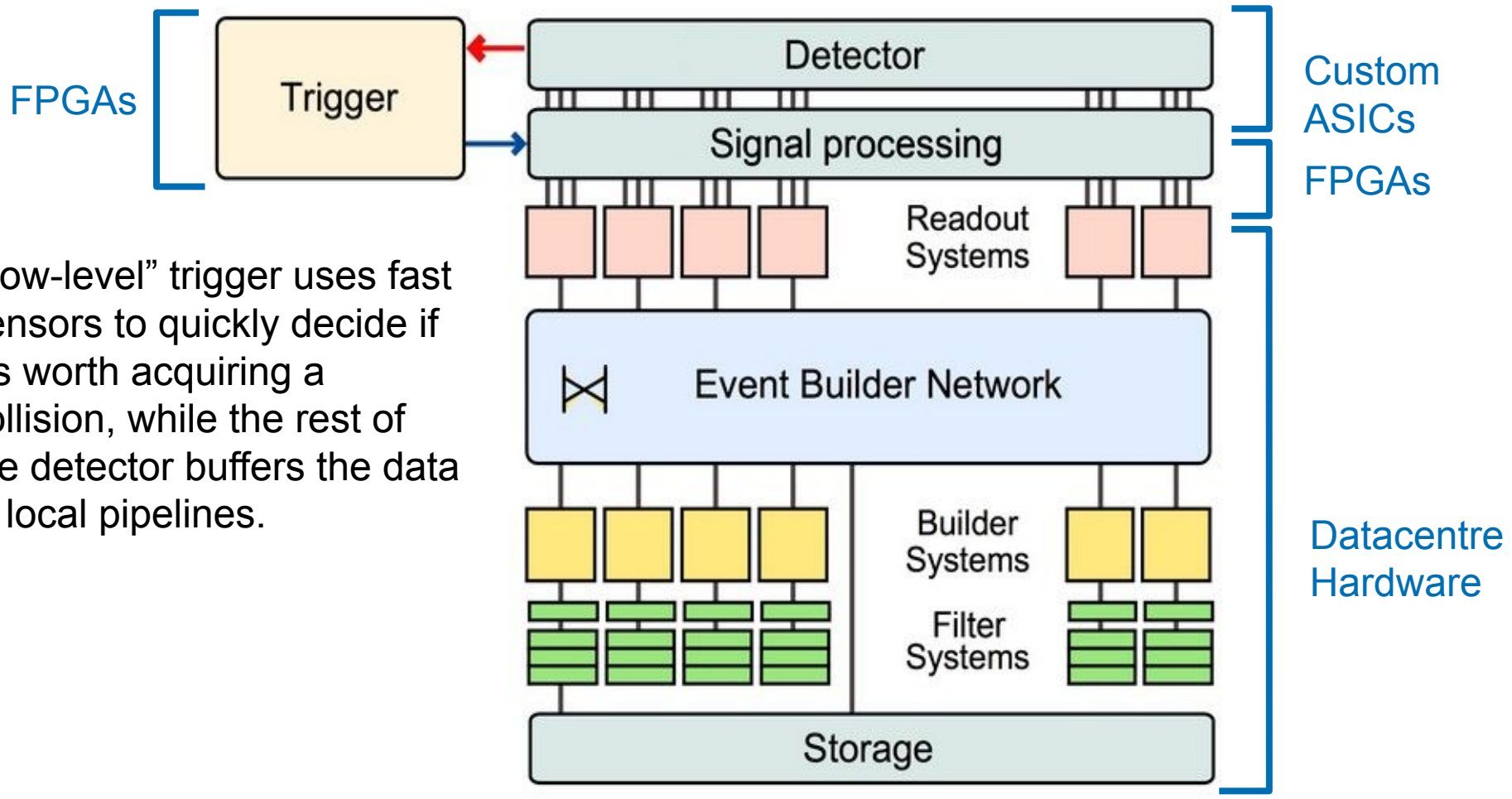
The haystack



This is where it hides: ~ 60 (LHC) – 200 (HL-LHC) other collisions producing 1000s of particles

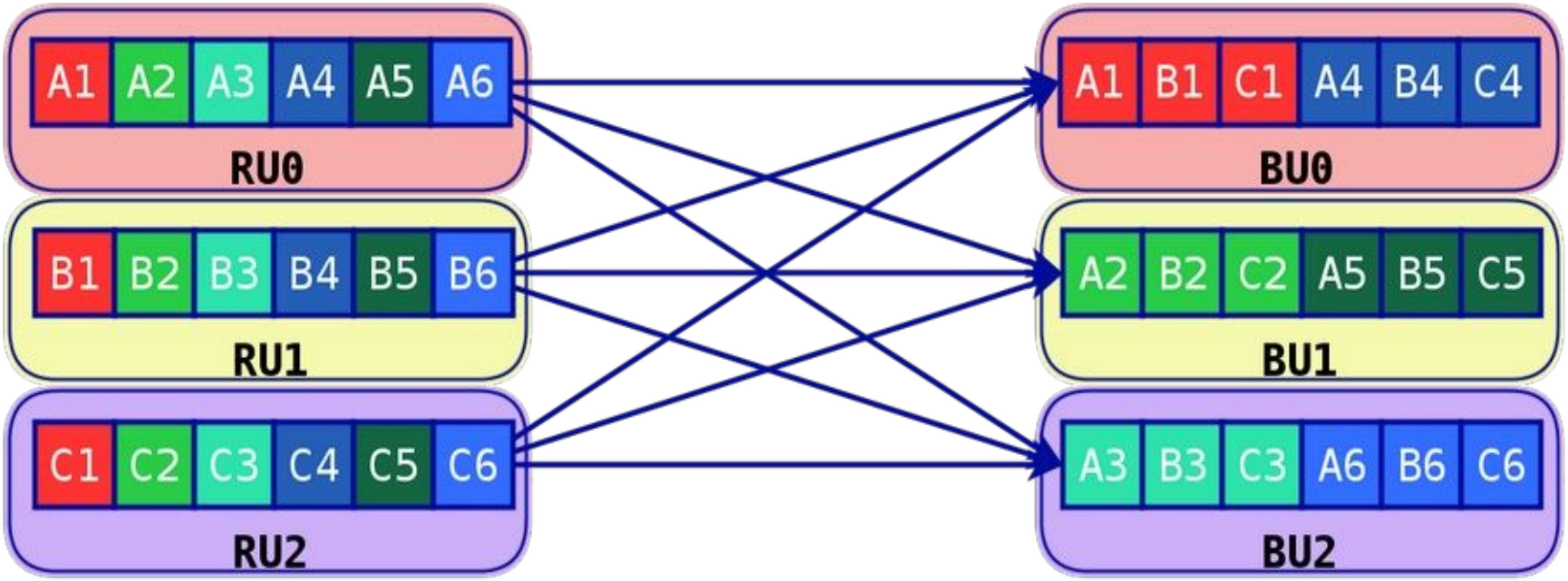
The LHC makes 40 million of these per second!

Data acquisition



"Low-level" trigger uses fast sensors to quickly decide if it's worth acquiring a collision, while the rest of the detector buffers the data in local pipelines.

Event building

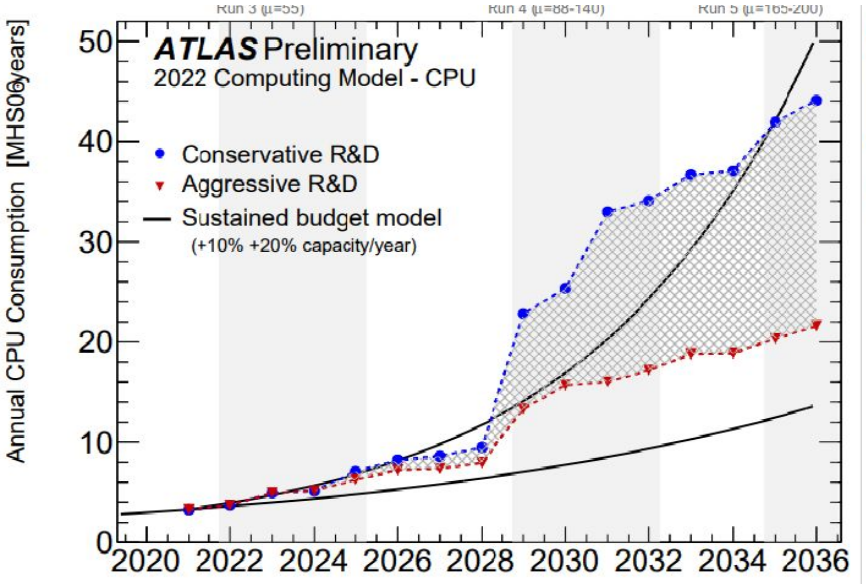


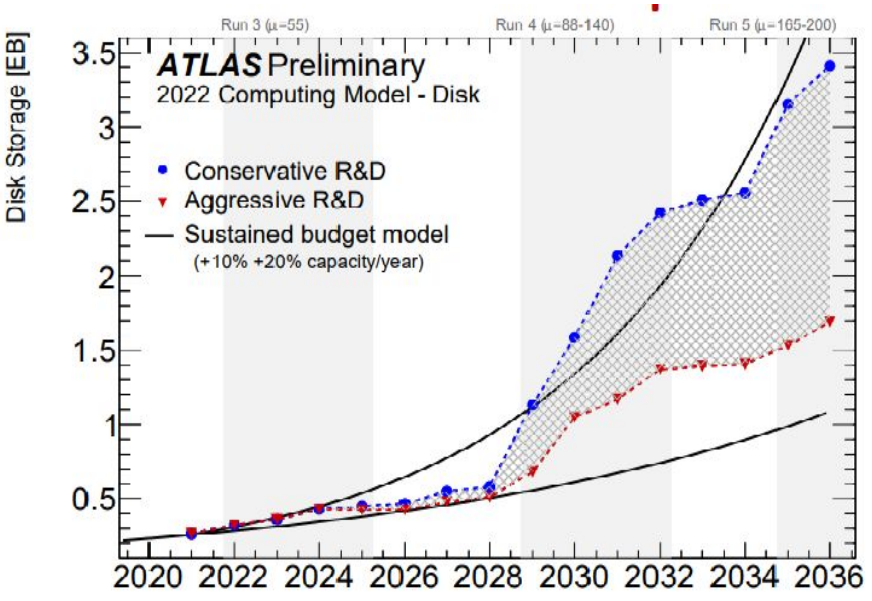
Readout unit (RU): receives processed signals from some sensors
Builder unit (BU): assembles all signals corresponding to the same observed phenomenon

Experiment predictions

ATLAS and CMS, the largest experiments, foreseen a ~10-15 times increase in the needs of CPU and Storage during HL-LHC compare to today.

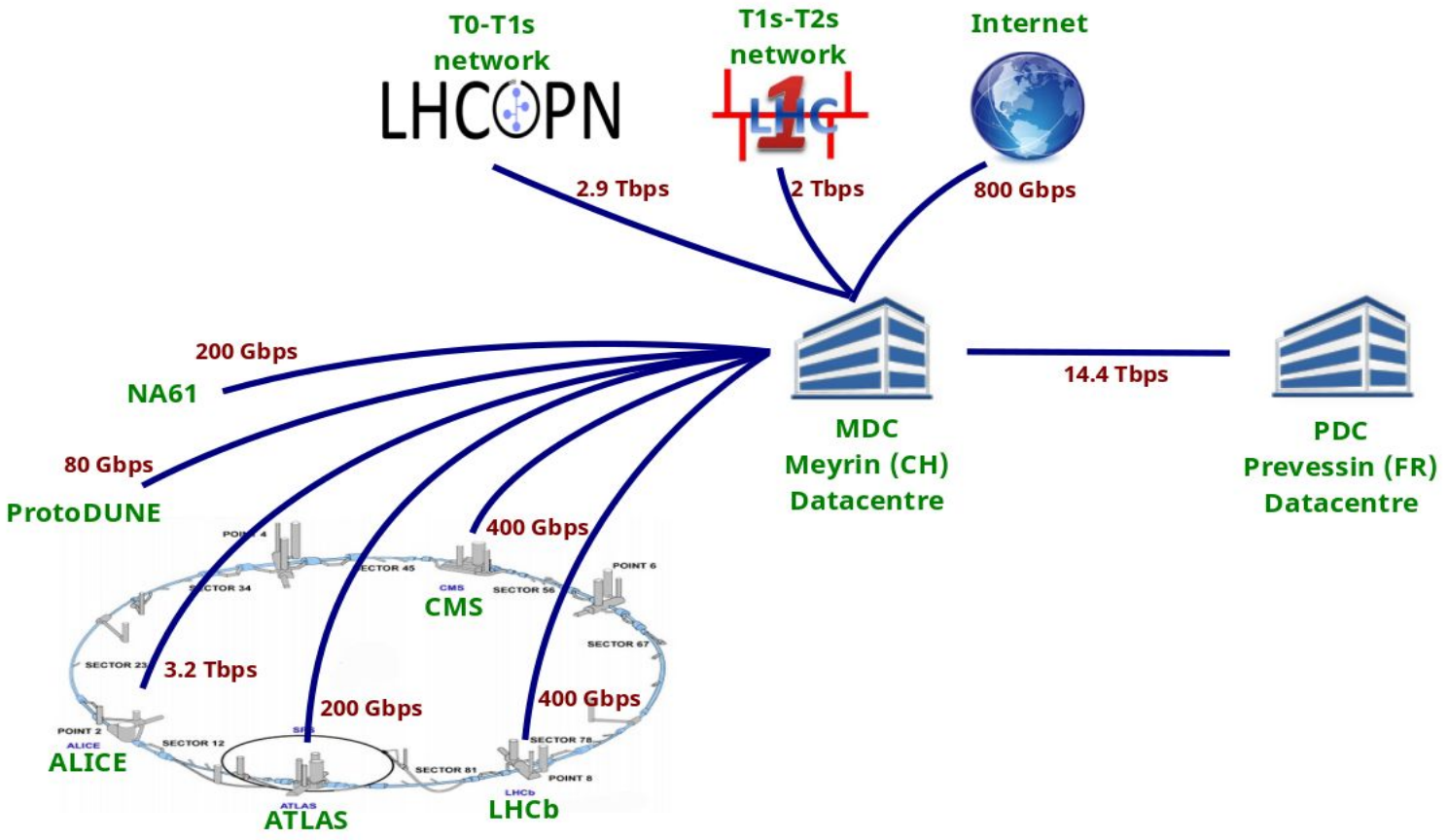
Network capacity will most likely have to grow by the same factor





Moving the Data

LHC data links

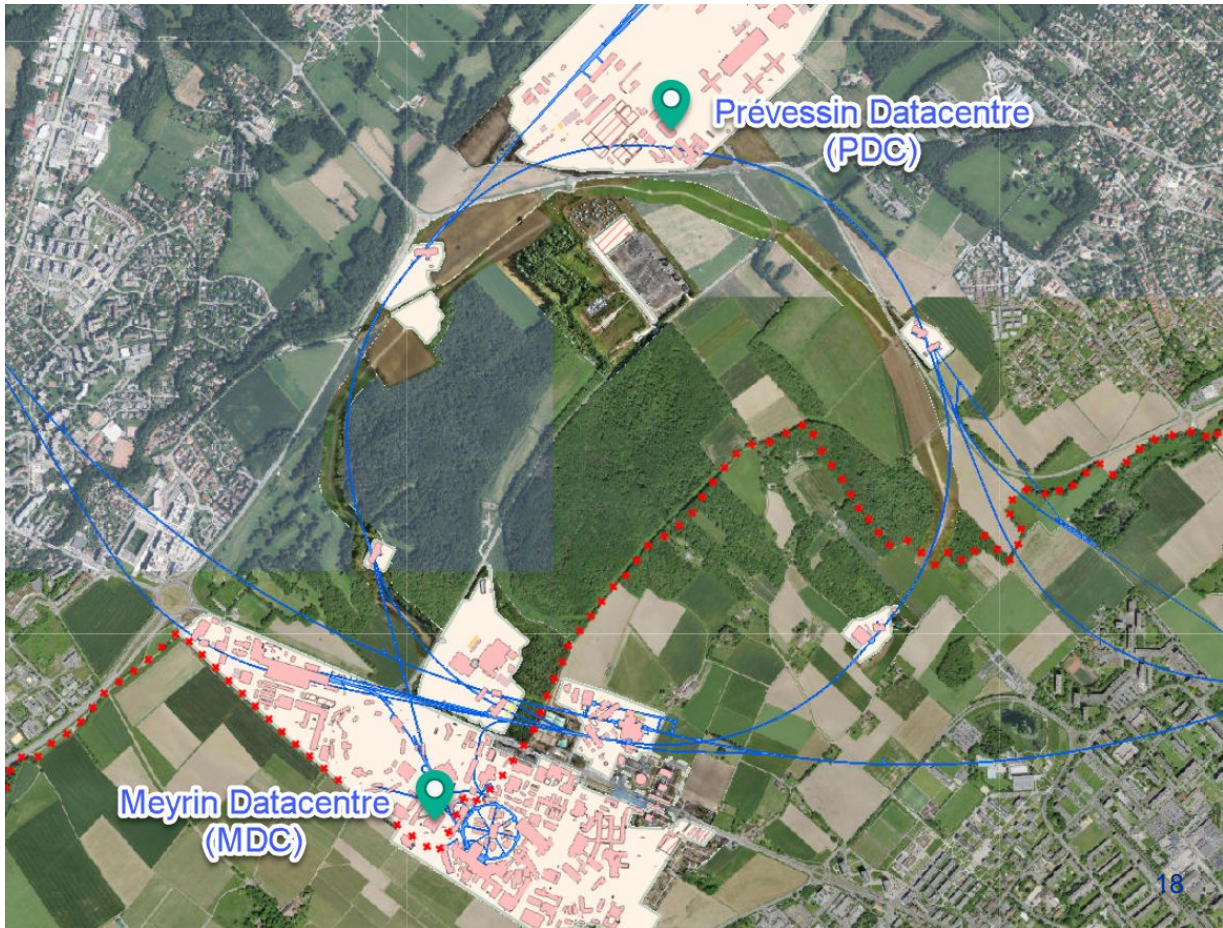


CERN data centres

Two locations:

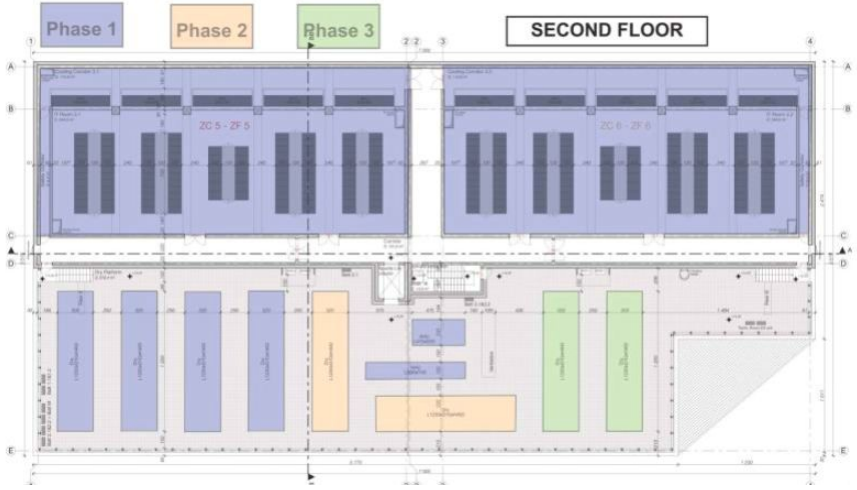
- Meyrin (CH) MDC
- Prevessin (FR) PDC

~ 5 km distance

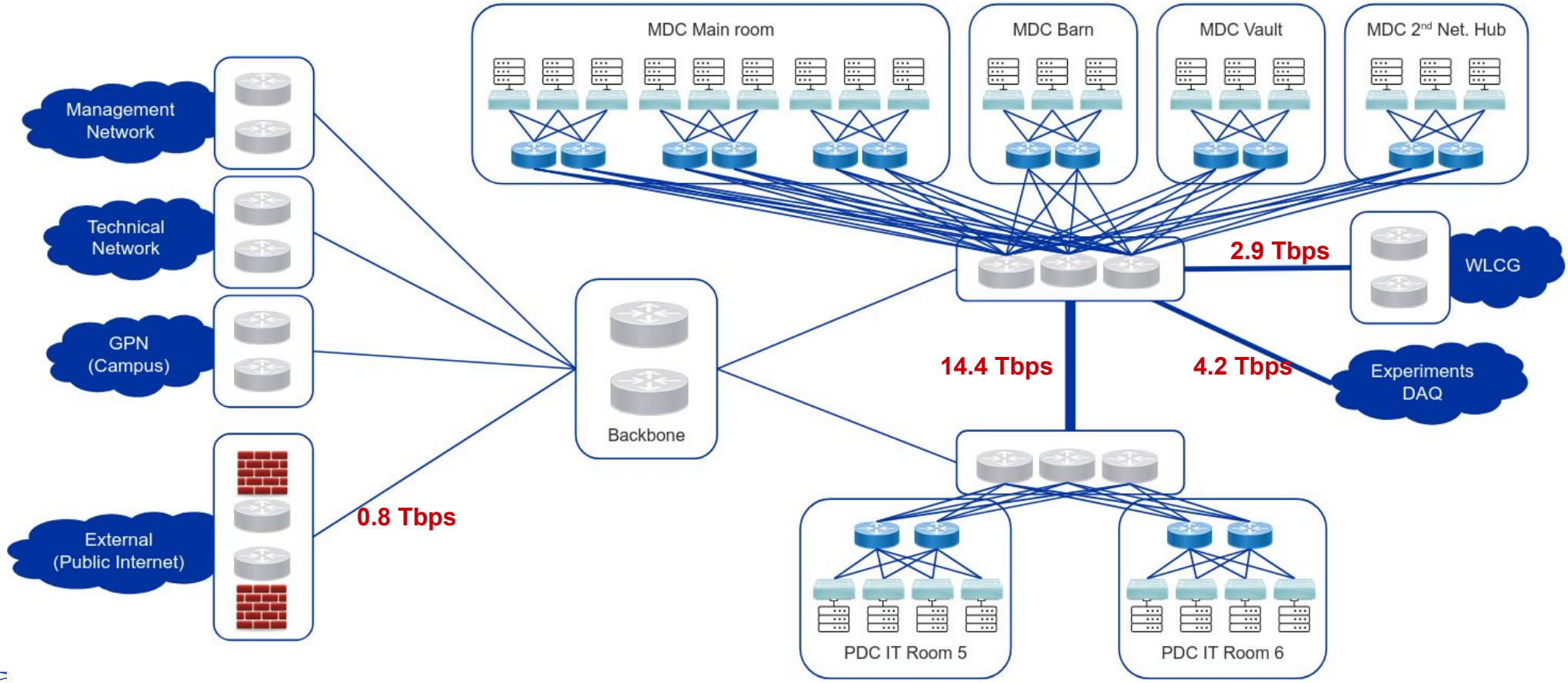


New Prevessin Data-Centre (PDC)

- **Construction completed**
- **Installation in 3 phases:**
  - 1st Phase: 4MW – 2nd floor (operational since 2024)
  - 2nd Phase: +4MW - 1st floor (~2027/2028 if approved)
  - 3rd Phase: +4MW - ground floor (~2030/2031 if approved)

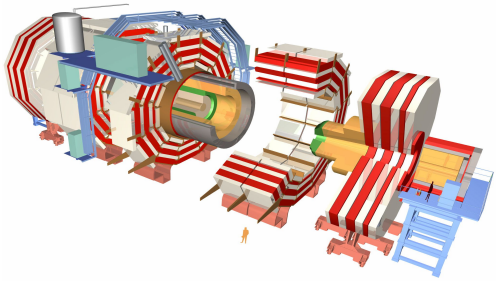


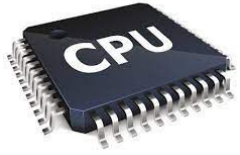
CERN data-centres network



Networks for WLCG

LHC Computing Model





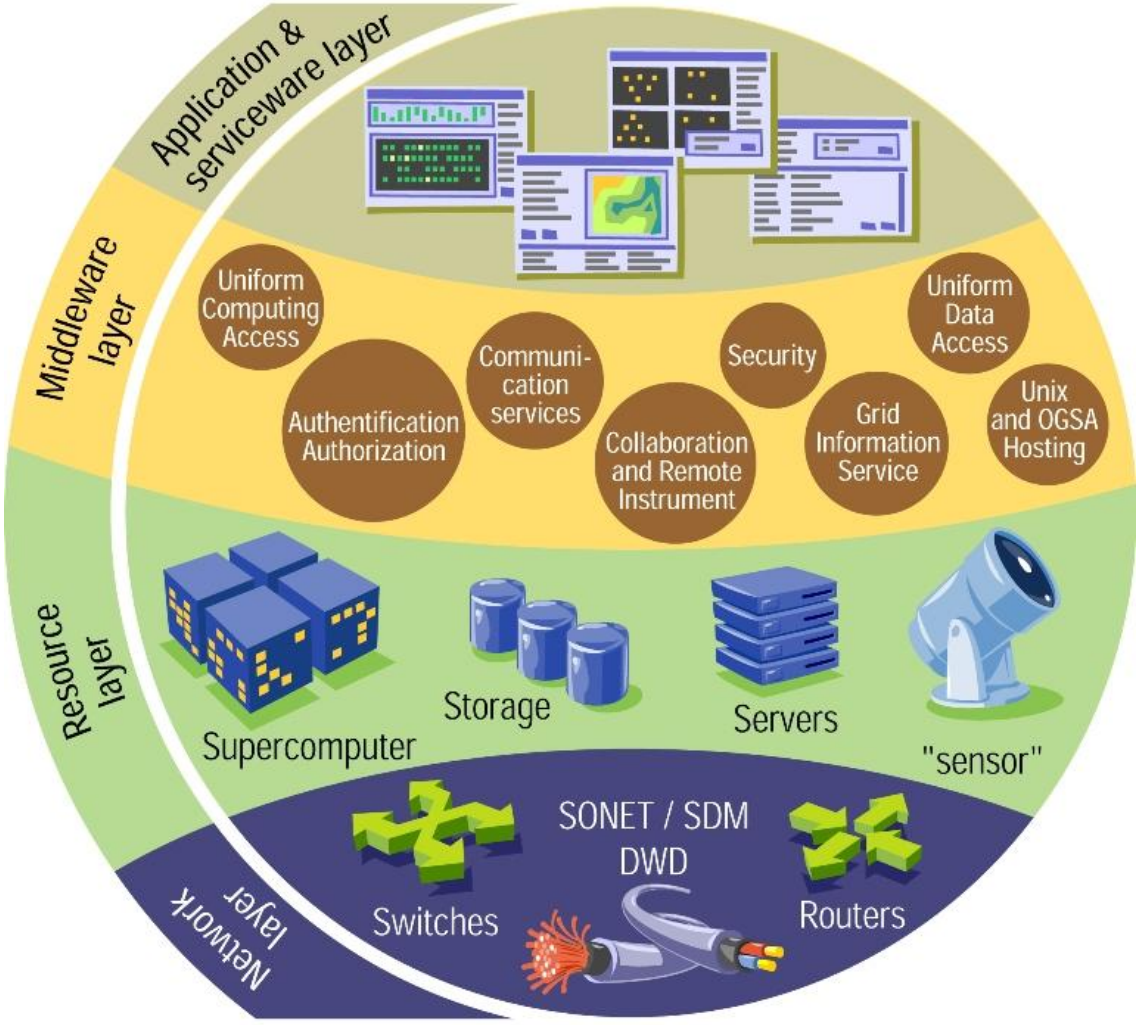
Tier 0 (1x)
Data source
Full data on Tape
Data reconstruction

Tier 1s (15x)
Distributed 2nd copy on Tape
Simulations, Data analyses

Tier 2s (~150x)
Data caches
Data analyses

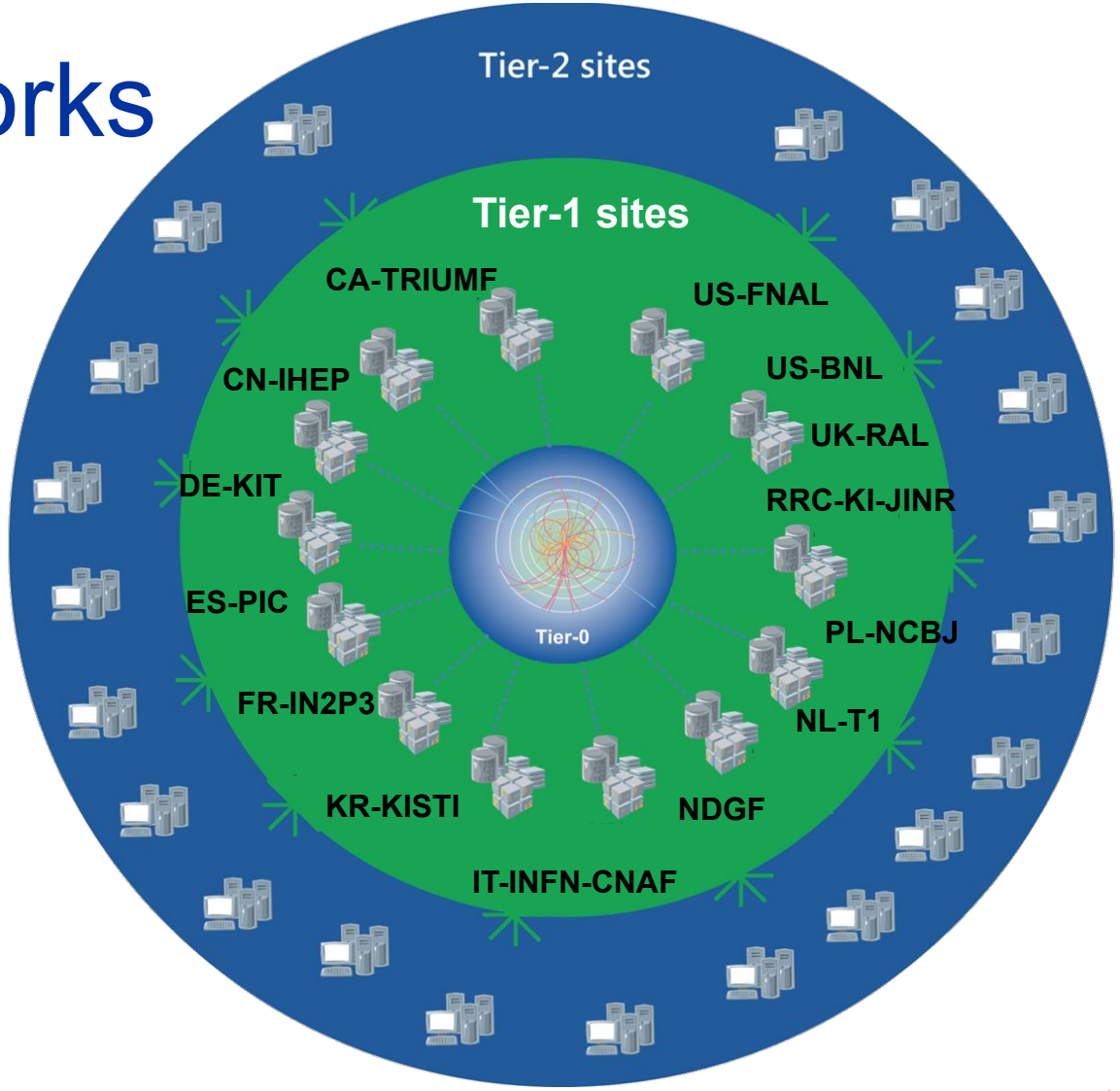
WLCG

The Worldwide LHC Computing Grid (WLCG) is **a large, distributed computing and storage infrastructure and the software framework to exploit it**

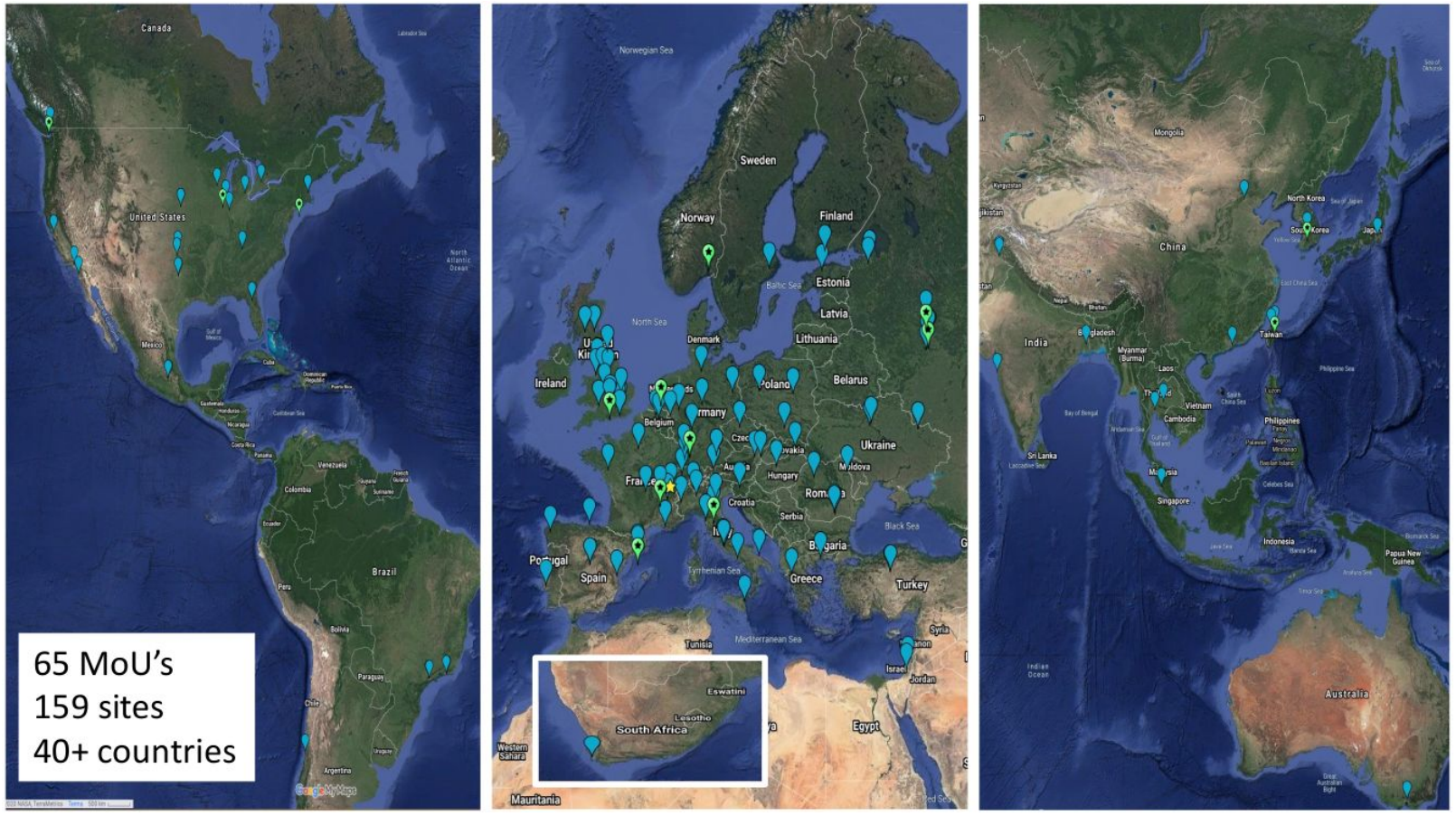


WLCG and networks

Computer Networks are an essential component of WLCG; they connect all the computing resources distributed in more than 150 institutes around the world



WLCG sites



HL-LHC network requirements for WLCG sites

Each Major Tier1s :

- 1 Tbps to the Tier0 (LHCOPN)
- 1 Tbps to the Tier2s (aggregated, LHCONE)

Each Major Tier2s:

- >400 Gbps (LHCONE)

WLCG and the NREN community are already working on the implementation of these requirements

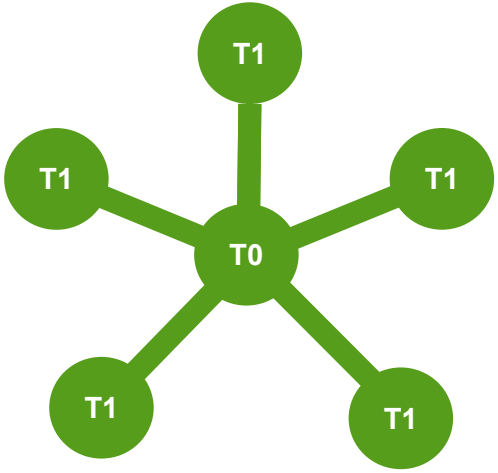
LHCOPN

LHCOPN

Private network connecting Tier1s to the Tier0

Secure:

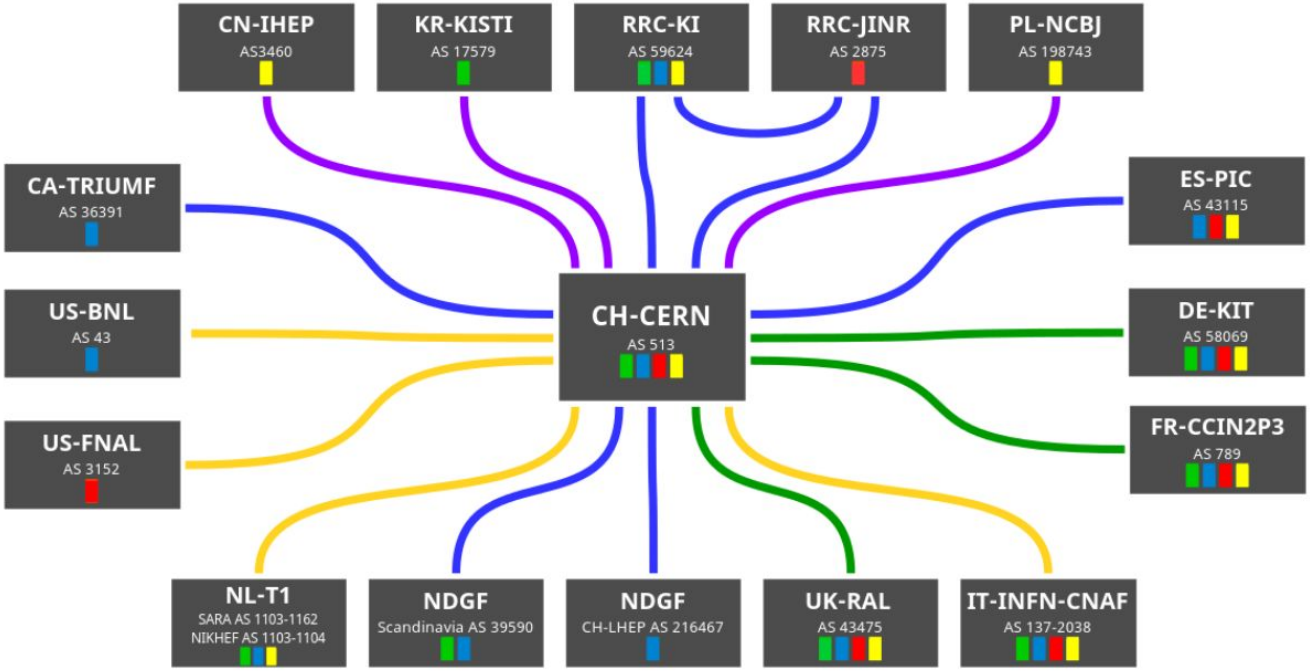
- Dedicated to LHC data transfers
- Only declared IP prefixes can exchange traffic
- Can connect directly to Science-DMZ, bypass perimeter firewalls

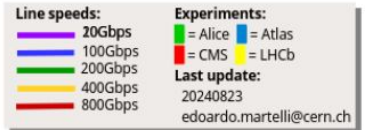


Technologies:

- L2 VPN
- BGP communities for traffic engineering

LHCOPN





https://twiki.cern.ch/twiki/bin/view/LHCOPN/OverallNetworkMaps

Numbers

- 17 sites for 15 Tier1s + 1 Tier0
- 14 countries in 3 continents
- 2.88 Tbps to the Tier0

LHCONE L3VPN service

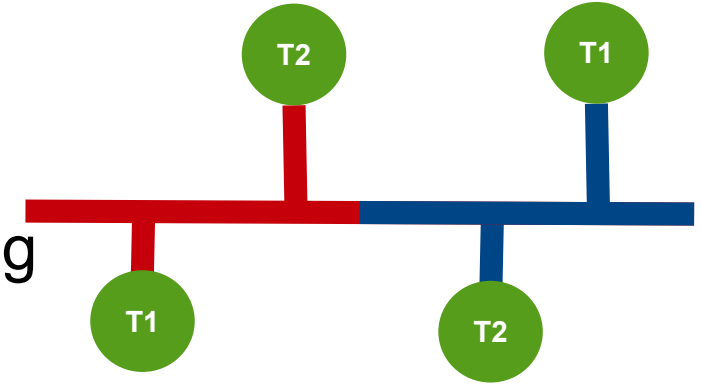
Private network connecting Tier1s and Tier2s

Secure:

- Dedicated to LHC data transfers
- Only declared IP prefixes can exchange traffic
- Can connect directly to Science-DMZ, bypass perimeter firewalls

Technologies:

- Multi domain L3 VPN
- BGP communities for traffic engineering

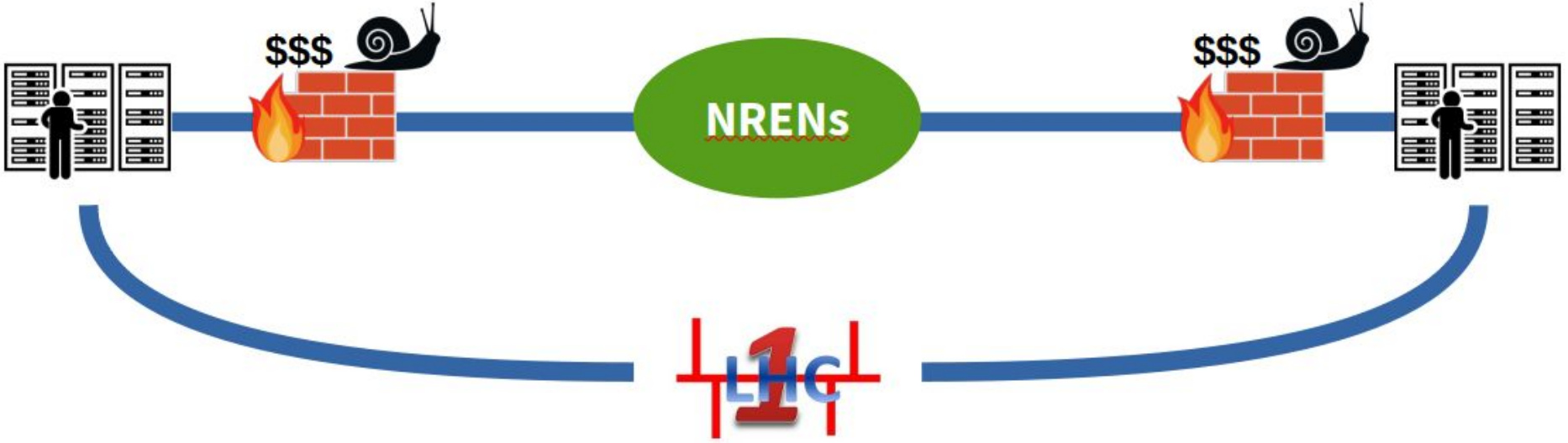


LHCONE status

- **VRFs: 32 national and international Research Networks**
- **Connected sites: ~110 in Europe, North and South America,Asia, Australia**
- Trans-Atlantic connectivity provided by ESnet, GEANT, Internet2, RedCLARA, NORDUnet, CANARIE and SURF
- Trans-Pacific connectivity provided by KREOnet, SINET, TransPAC
- Interconnections at Open Exchange Points including NetherLight, StarLight, MANLAN, WIX, CERNlight, Hong Kong, Singapore and others

Why LHCONE is useful

LHCONE is an overlay network, it doesn't bring more bandwidth by itself.
But it is a trusted network, more secure than a generic Internet upstream. Thus LHCONE can be connect directly to the data-centres and bypass low-bandwidth/expensive perimeter firewalls



WLCG Data Challenges

HL-LHC network requirements

ATLAS & CMS T0 to T1 per experiment

- 350PB raw data per year; average of 50GB/s or 400Gbps during LHC running time
- Another 100Gbps estimated for prompt reconstruction data tiers (AOD, other derived output)
- estimated 1Tbps for CMS and ATLAS summed

ALICE & LHCb T0 Export

- 100 Gbps per experiment estimated from Run-3 rates

Minimal Model

- Sum (ATLAS,ALICE,CMS,LHCb)*2(for bursts)*2(safety-margin) =

  4.8Tbps expected HL-LHC bandwidth

Flexible Model

- Experiments may need to reprocess and reconstruct the collected data during the year
- This requires doubling the bandwidth of the Minimal model:

  9.6Tbps expected HL-LHC bandwidth

Data Challenges for HL-LHC

WLCG organises a series of data challenges to progressively prepare for HL-LHC data taking

- Demonstrate readiness for the expected HL-LHC data rates with:
  - Increasing volume/rates
  - Increase complexity (e.g. additional technology)

2021: 10% of HL-LHC requirements *(480Gbps minimal – 960Gbps flexible)*

2024: 25% of HL-LHC requirements *(1.2Tbps minimal – 2.4Tbps flexible)*

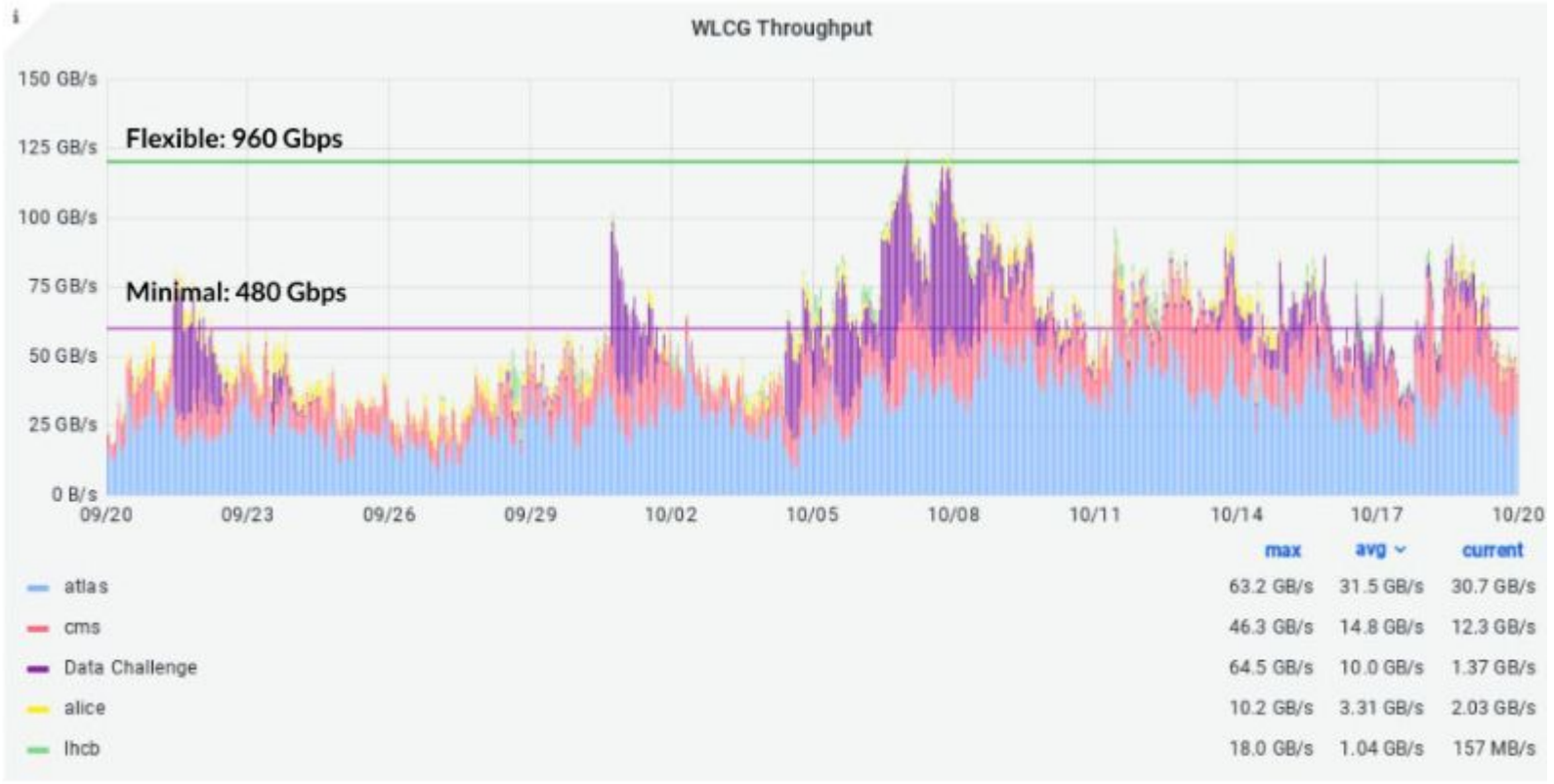
2027: 50% of HL-LHC requirements *(date and % to be confirmed)*

2029: 100% of HL-LHC requirements *(date and % to be confirmed)*

2030: start of HL-LHC (Run4) *(4.8Tbps minimal – 9.6Tbps flexible)*

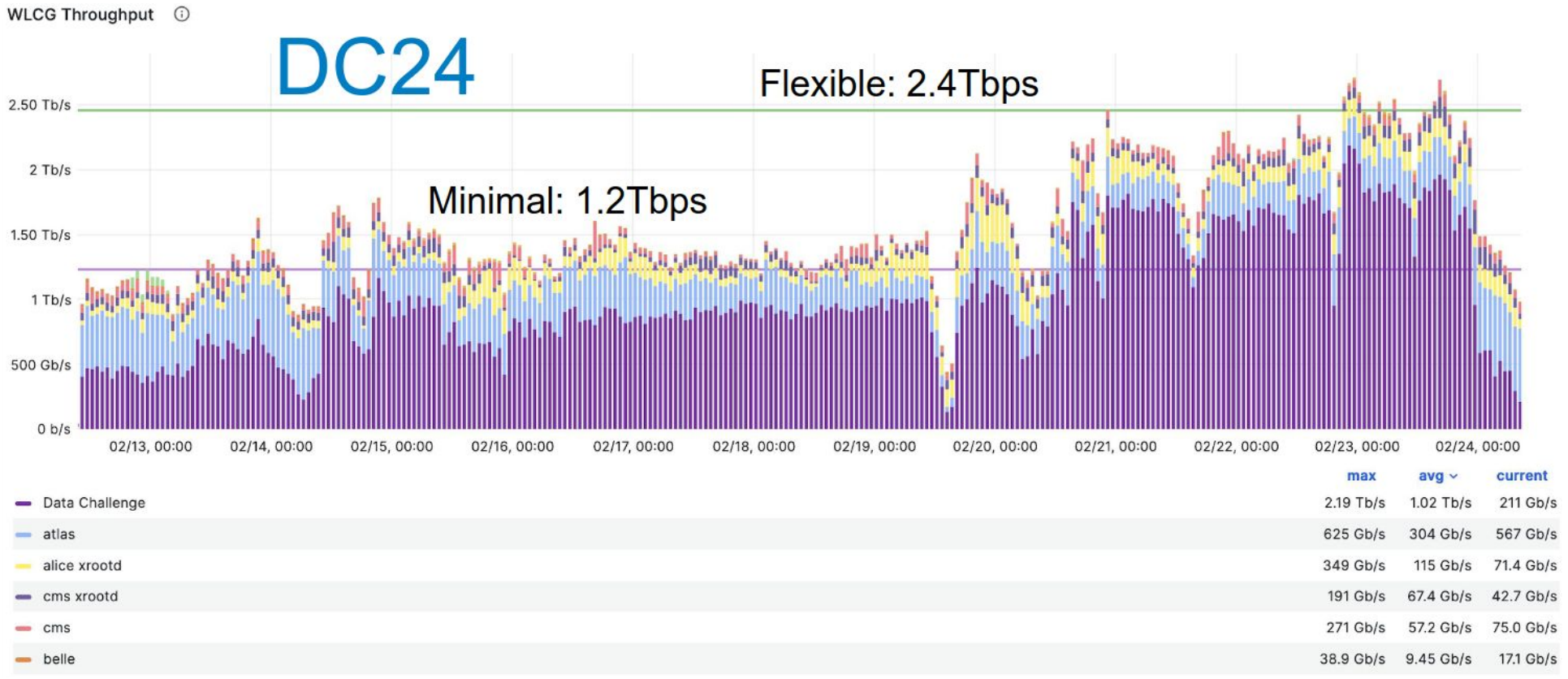
DC21

Flexible target reached for a short time



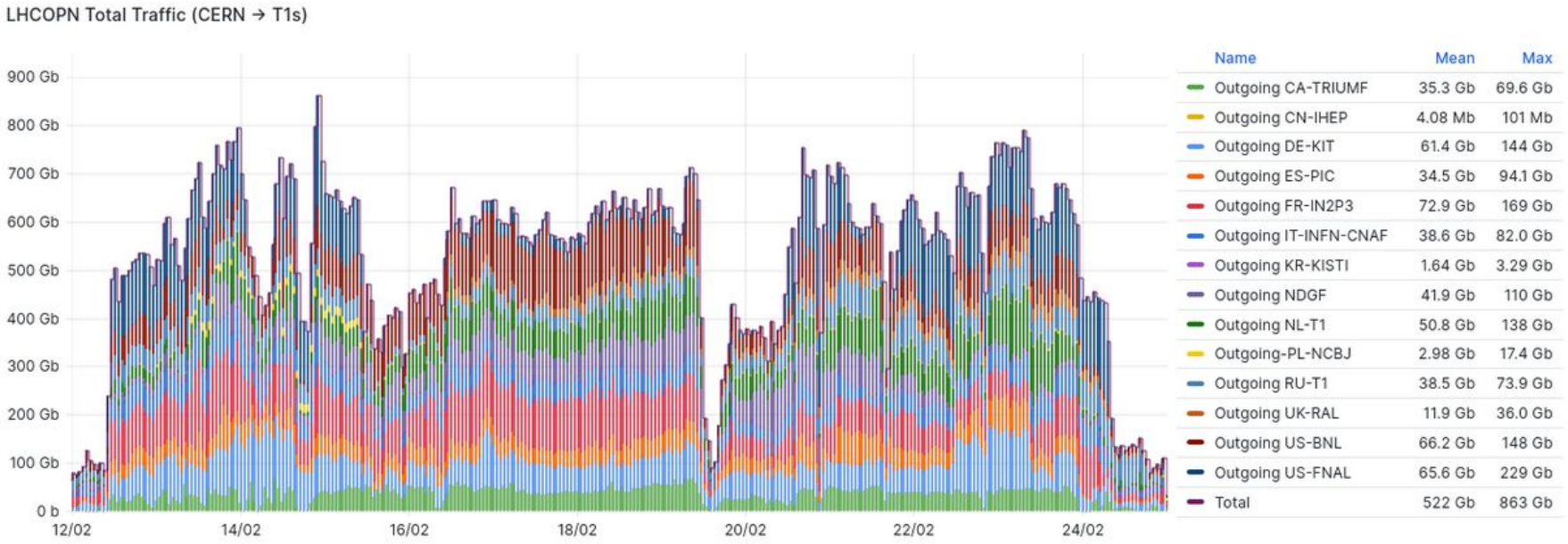
DC24

Achieved full throughput of minimal model. Flexible only for short time.



LHCOPN and DC24

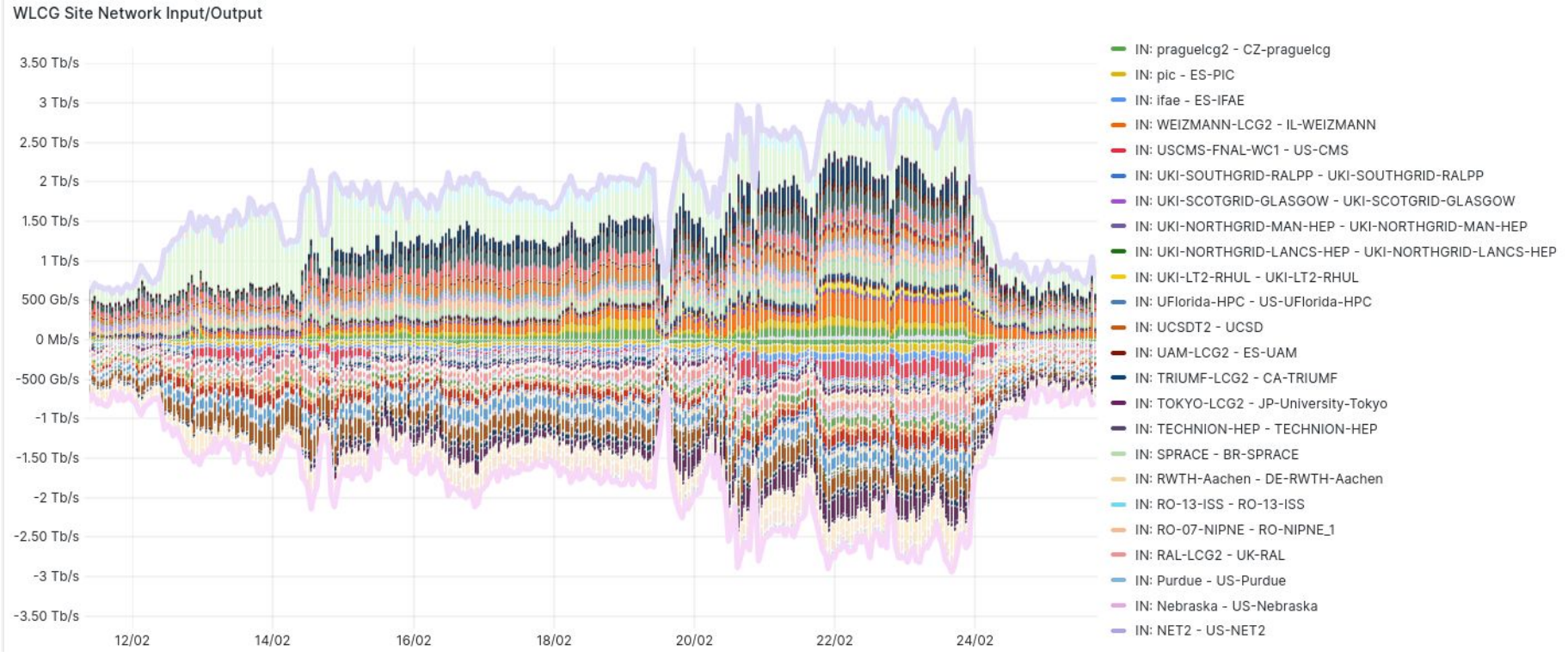
Tier0-Tier1 traffic on LHCOPN: peak at 800Gbps
- used ~35% of existing bandwidth (2,6Tbps aggregated)
- required to have >10Tbps for HL-LHC



https://monit-grafana-open.cern.ch/d/HreVOyc7z/all-lhcopn-traffic?orgId=16&var-source=long_term&var-bin=1h&from=1707650640915&to=1708873382950

DC24 site monitoring

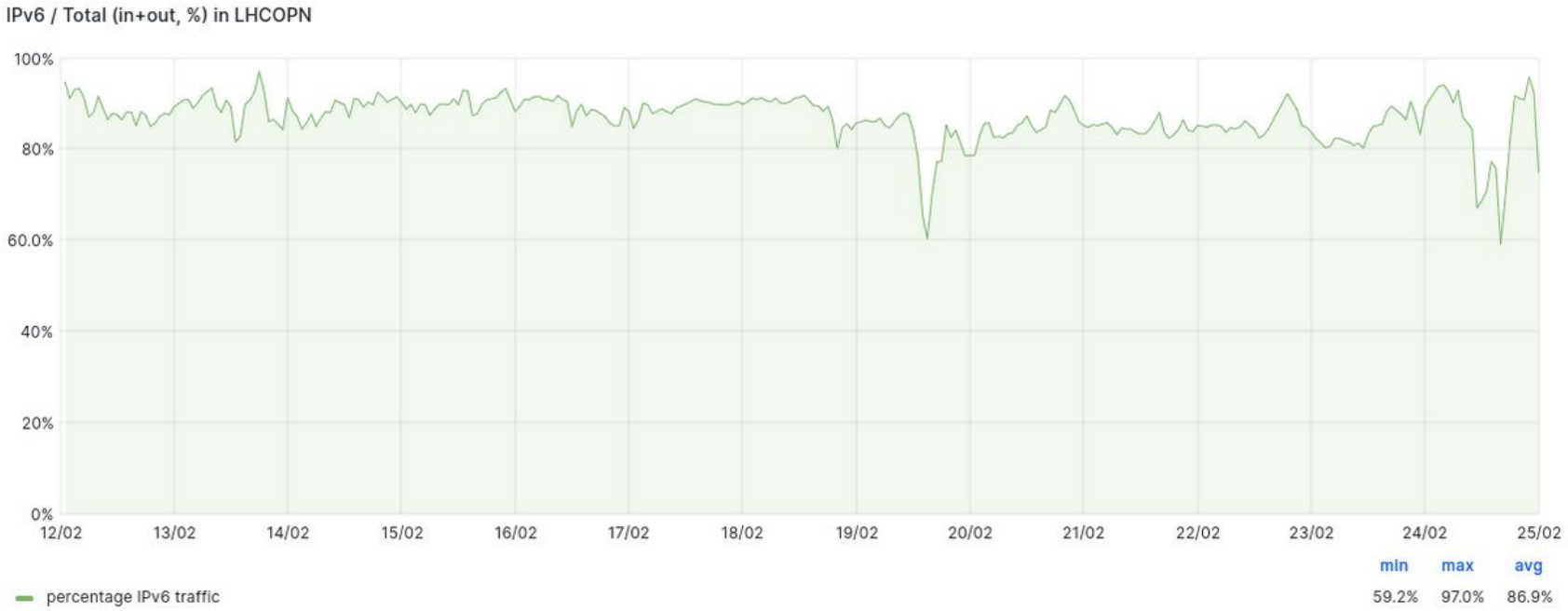
WLCG aggregated traffic exceeded 3Tbps (partial view)



https://monit-grafana-open.cern.ch/d/MwuxgogIk/wlcg-site-network?from=1707638857216&orgId=16&to=1708880381142

IPv6 and DC24

- IPv6 Traffic in LHCOPN: 86.9% of the total
- Some sites testing IPv6 because of IPv4 scarcity



DC24 results on Networking

- Research & Educations (REN) networks demonstrated more than sufficient capacity and reliability during DC24 and were NOT a bottleneck for any of the experiments

- Some sites did identify local network bottlenecks or non-optimal architectures

- Various network technologies (NOTED, SENSE, BBR, perfSONAR, SciTags, Spectrum sharing…) were successfully tested during DC24 and showed promising results, motivating the effort to put them into production.

- Although R&E networks were not a bottleneck for DC24, storage infrastructure and middleware are being improved and R&E networks need to keep pace

- We will need regular mini-challenges to track progress and prepare for DC27

Plans for future Data Challenges

DC27

- 50% HL-LHC (~Feb 2027)
- 25% of the ‘flexible’ target was already hugely challenging. Is 50% too ambitious?
- Still 3 years before the start of HL-LHC (sites may not have HL-LHC hardware in place, or be paying for HL-LHC network connection)
- New FTS version? Tokens? Tapes?

Mini data challenges

- Keep up momentum in the 3 year gap between DC24 and DC27
- Continue to try to improve the existing infrastructure
- Capacity and Capability mini challenges are in progress

Challenges very useful to Network Providers to show the procured resources are in line with real network utilisation. Also useful to highlight network bottlenecks and work with sites to remove them.

Network R&D

WLCG guidelines

In the next 10 years WLCG Networking will be faced with two major challenges:

- dealing with the HL-LHC data volumes and complexity
- cohabitation with other experiments and sciences on the same infrastructure

WLCG, together with the R&E network community, needs to play a leading role:

- **modernise network services**, progressing with the ongoing R&D activities and bringing early prototypes in production
- **engage with other experiments and sciences** to drive the evolution of R&E networks

Network Requirements for the next decade

From WLCG and the LHC experiments:

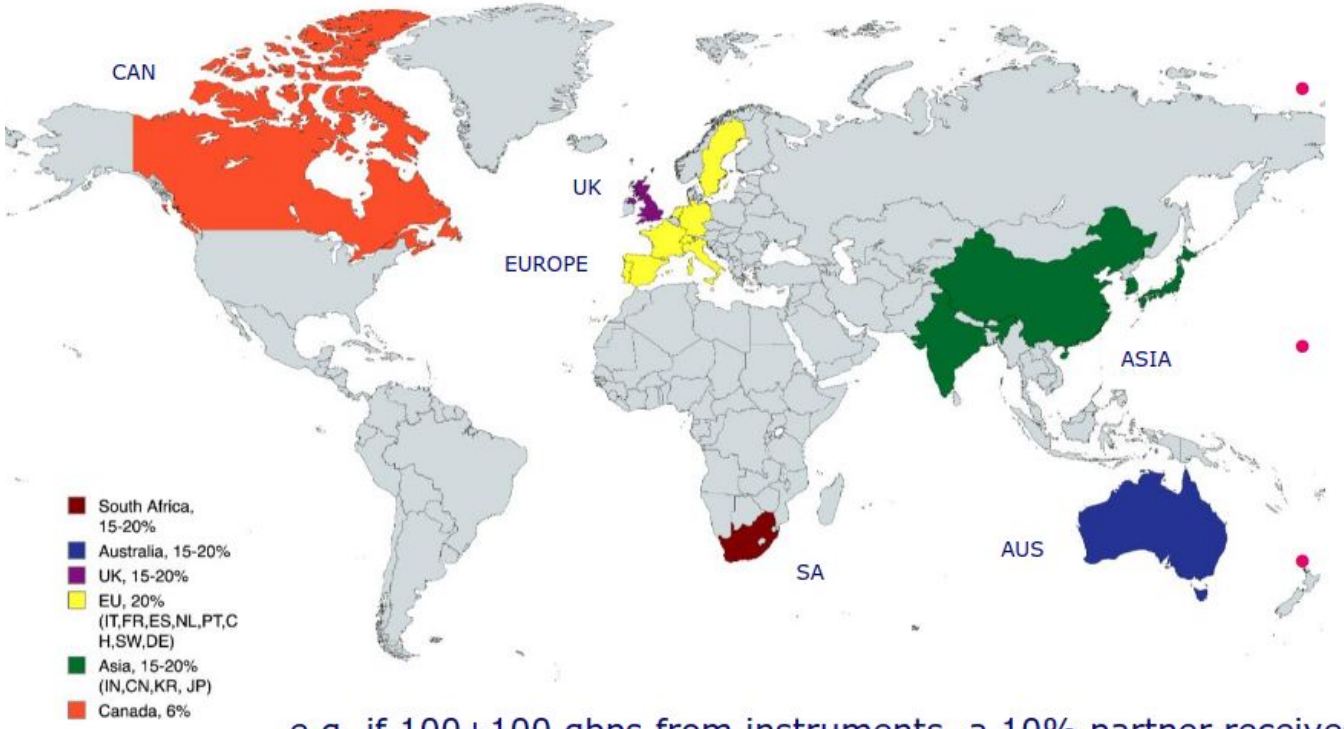
- Enough bandwidth to cope with LHC and other large science projects needs
- Visibility of network health status
- Visibility of network utilization
- Predictability of network performances
- Security at Terabit scale

Co-existence with other large data science projects

At the time of HL-LHC (2030), other large data science projects will come on-line

It will be important for R&D network providers to be aware of all requirements and data flows to collectively plan the necessary upgrades

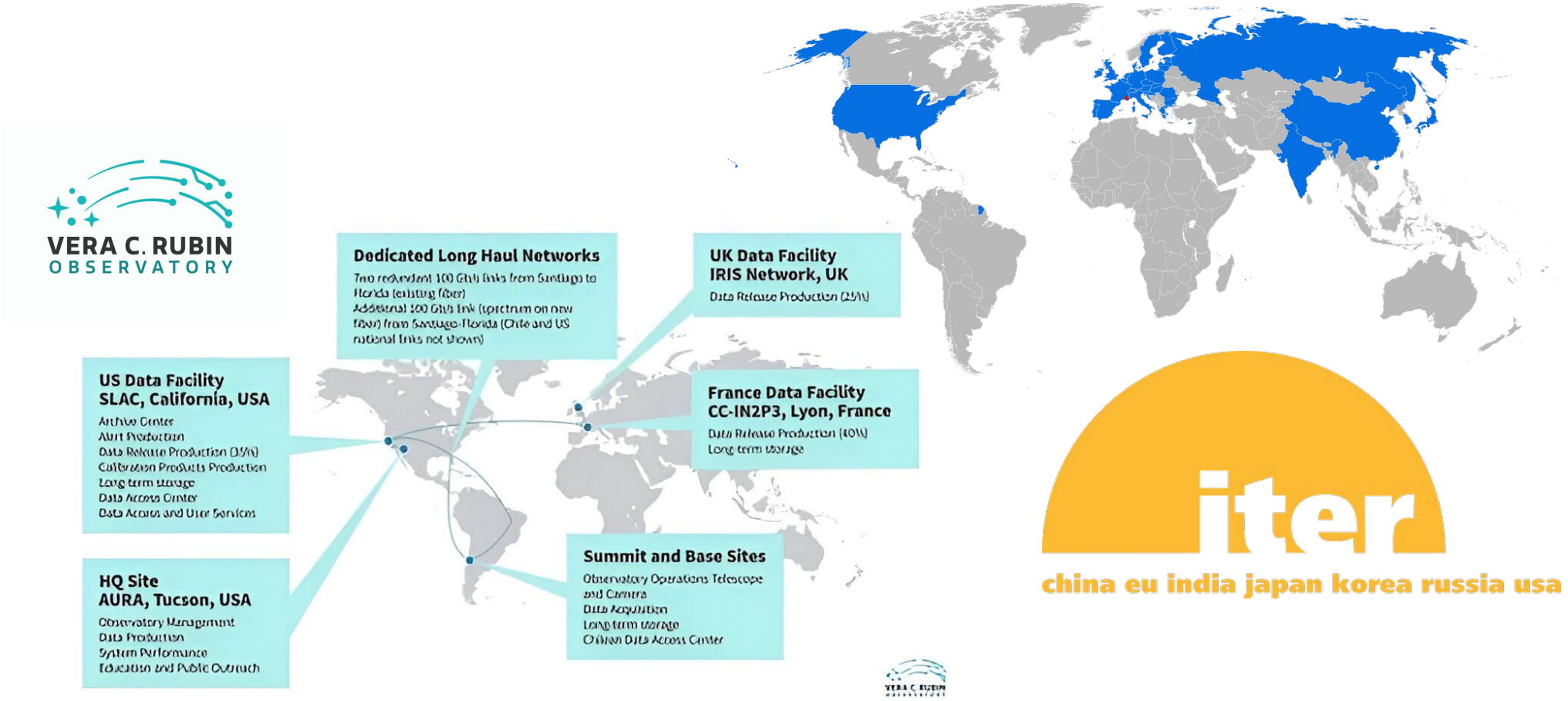
SKAO network provisioning



- Roughly, 6 global zones of equivalent size (Canada smaller)
- **Distribute two base copies** of each data product to different countries, and perhaps insist to different regions
- Average incoming rate per (20%) region not more than 2x40 Gbit/s = 80Gbit/s (~2x12 Gbit/s for Canada)
- **Modelling assumes average 100 Gbit/s out of SA and AUS**

e.g. if 100+100 gbps from instruments, a 10% partner receives 40gbps data (400 TBytes per day, 140 PBytes per year)

More Big Data sciences coming on line



Network visibility

The LHC experiments have been asking for more visibility of network utilization and performance, to better understand the impact of their applications and to improve their efficiency

perfSONAR

perfSONAR network monitoring platform

- Developed by the collaboration of Internet2, GEANT, ESnet, RNP, Indiana University, University of Michigan
- Toolkits installed at NRENs PoPs and WLCG sites
- Essential to monitor WLCG network performances and investigate issues

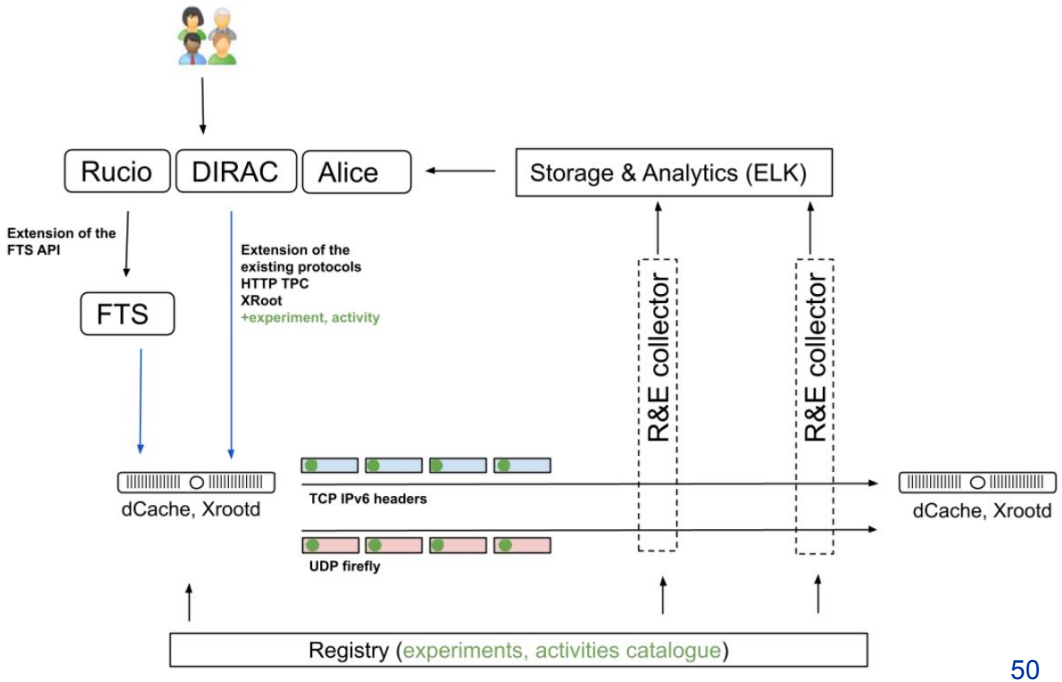
perf2ONAR
years

SciTags

Marking of data packets and flows with Experiment and Application IDs for better accounting

Two options being pursued:

- Tag in the IPv6 flowlabel field
- Tag (and more) in UDP fireflies (UDP packets sent in parallel to each flow)



Predictability of network utilization

Scientists generating and manipulating very large amount of data, sometime get worried that the network may become a scarce resource, congested by their data transfers, transfers which may then take forever to complete.

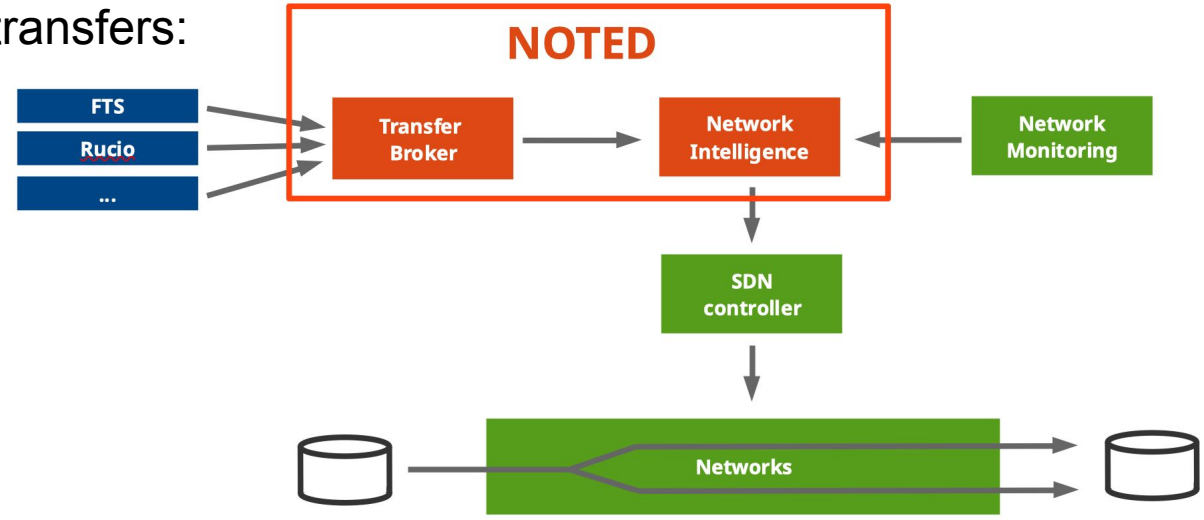
Several projects have been developed to allow applications to increase network efficiency

NOTED SDN

NOTED is a framework that can interface with File Transfer applications to detect large data transfers, src-dst and their duration. Then it can trigger network optimization actions to speed up the execution of those transfers

Already used with production data transfers:

- During SC22 and SC23
- New version with triggers from Network Monitoring tested during DC24



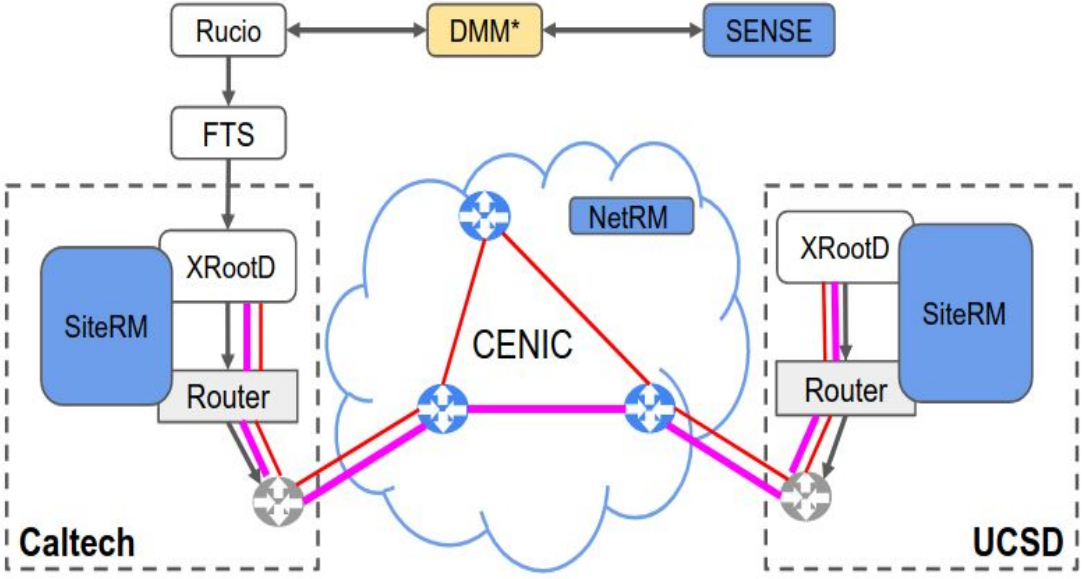
- SC: Super Computing
- DC: (WLCG) Data Challenge

Using SENSE to move CMS data in Rucio

Project led by UCSD and Caltech

Objectives of the project:
- #1 Make Rucio capable to schedule transfers on the network and prioritize them
- #2 Predetermined transfer speed and quality of service

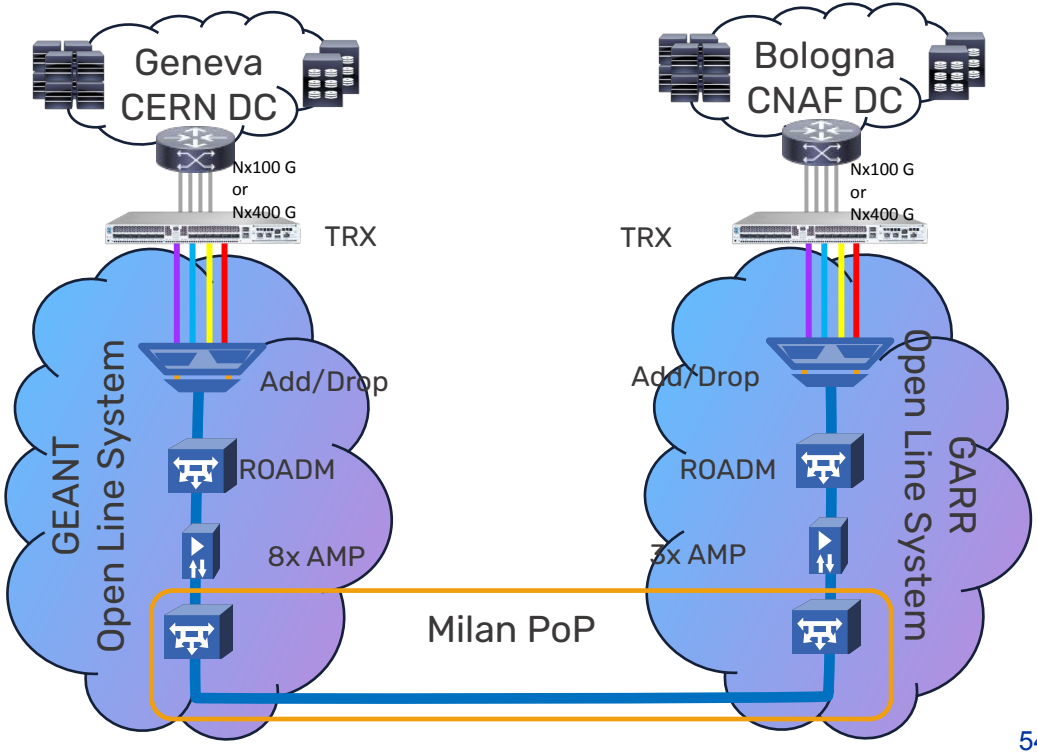
Demonstrated:
- SENSE can build VPNs between pairs of XrootD servers in charge of FTS transfers requested by Rucio
- QoS can be provisioned in the network to prioritize the traffic in the VPN



CNAF-CERN DCI

The LHCOPN link of IT-INFN-CNAF Tier1 is implemented using shared spectrum over GEANT and GARR (Italian NREN) dark fibres

- 4x100Gbps links between CERN and CNAF used for DC24 and now in production
- Ready to be upgraded to 4x400Gbps
- cost effective technique to get >1Tbps LHCOPN connections already today



Security at Terabit scale

Cyber attacks are the needle in a haystack of data transfer flows. Security will be more and more challenging for scientific data-centre.

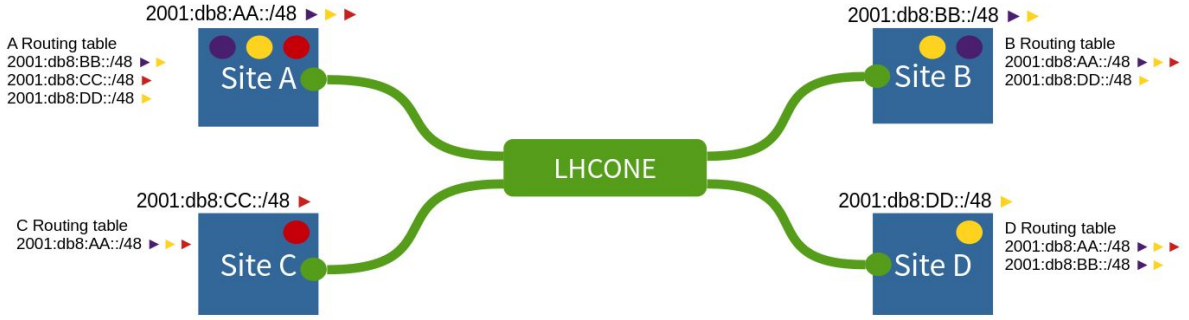
How R&E networks can help?

MultiONE, or LHCONE prefix tagging

LHCONE success and grow may undermine its major value: trust

Sites have been requested to tag their LHCONE prefixes with the ID of the experiment using it

Goal: reduce data-centre exposure



Conclusions

Summary

- HL-LHC will increase data production of a factor of 10. Networks will have to increase capacity while keeping a sustainable cost

- WLCG and R&E networks are preparing for the new accelerator. Data Challenges are helping software and networks to meet the HL-LHC requirements

- The collaboration between R&E networks and WLCG is crucial for the development and testing of new network technologies

Credits

With contributions from:
- Katy Ellis (RAL)
- Mario Lassnig (CERN)
- Christoph Wissing (DESY)
- Shawn McKee (University of Michigan)
- Stefano Zani (INFN-CNAF)
- Bernd Panzer (CERN)
- Tony Cass (CERN)
- Tommaso Colombo (CERN, LHCb)
- Vincent Ducret (CERN)

Questions?