

Network activities update

ALICE workshop

4th of May 2017

edoardo.martelli@cern.ch



Agenda

Updates on:

- Networks at CERN IT
- LHCOPN
- LHCONE
- IPv6 adoption
- Cloud Activities [if time allows]

CERN IT

IT Communication Systems

The IT-CS group is responsible for communication services in use at CERN for data, control, voice and video

<http://information-technology.web.cern.ch/about/organisation/communication-systems>



Recent activities

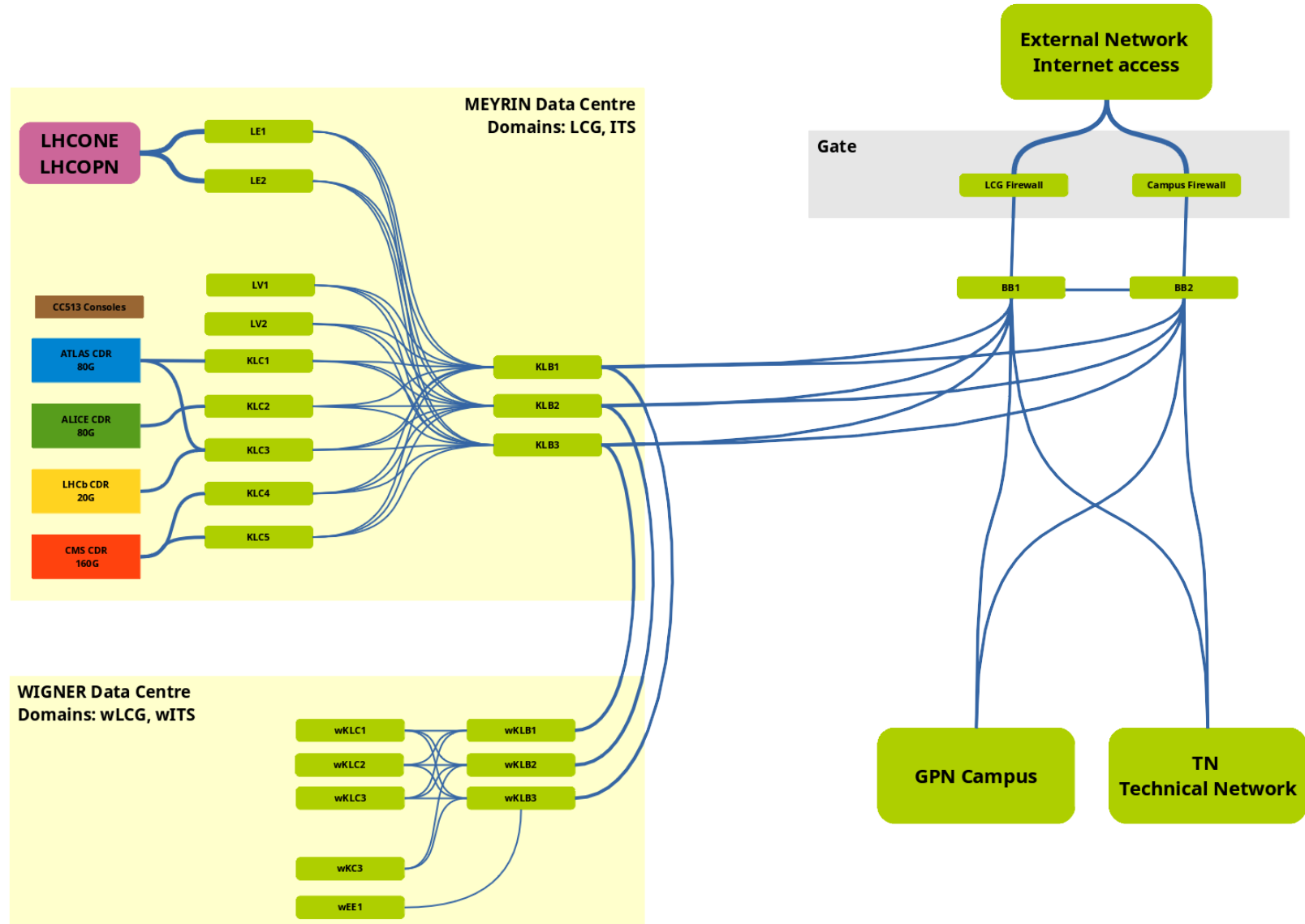
Wigner remote data-centre:

- added 3rd 100G link Geneva-Budapest
- re-configured network to increase local capacity

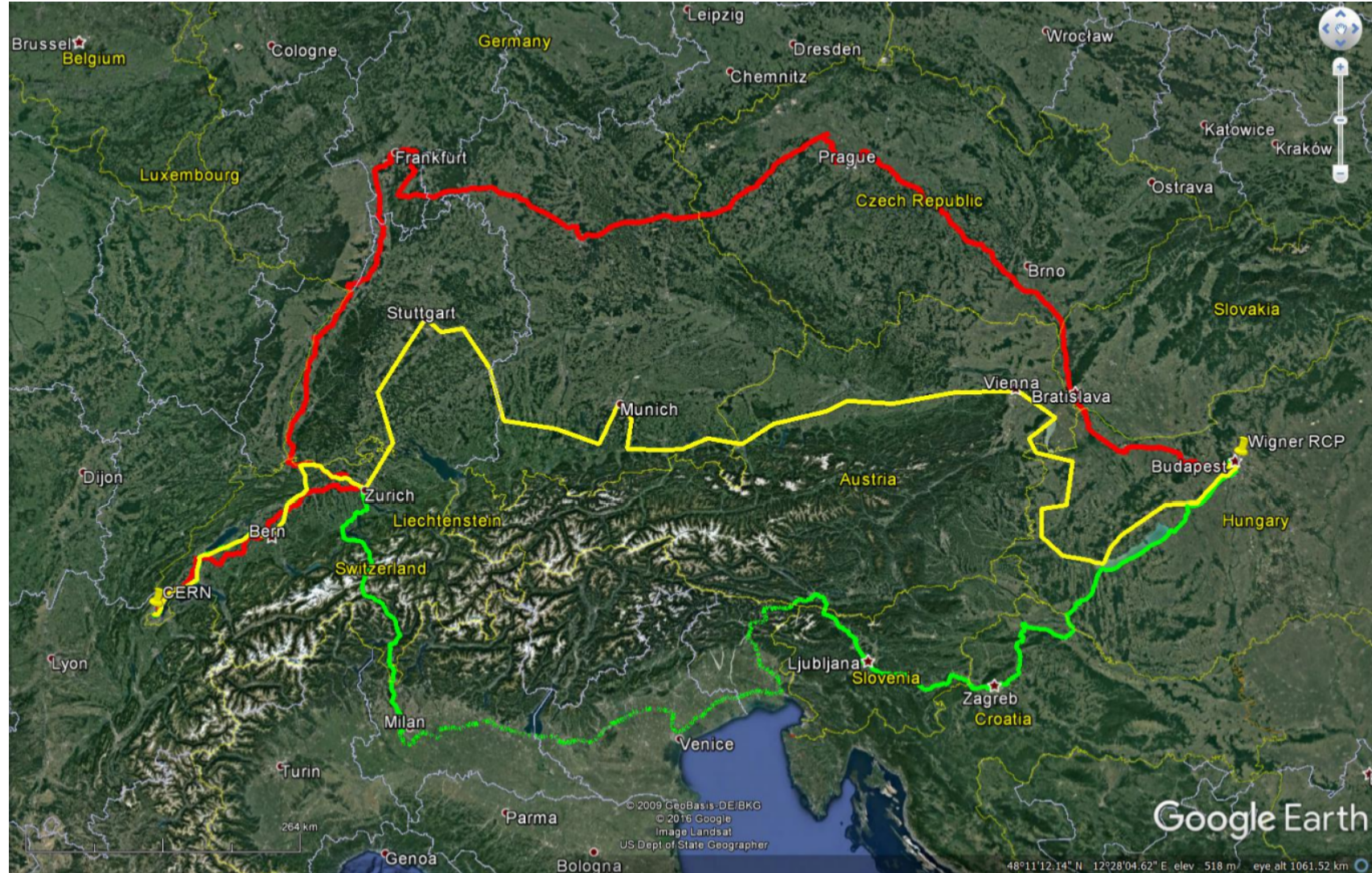
Tender for new datacentre routers

- to replace Brocade MLXE contract
- on going, waiting for replies
- to be concluded by September

CERN data centres – current setup



3rd 100G Geneva-Budapest link



Recent activities

Campus upgrades

- New WIFI network being deployed
- Planning for routers and switches upgrade

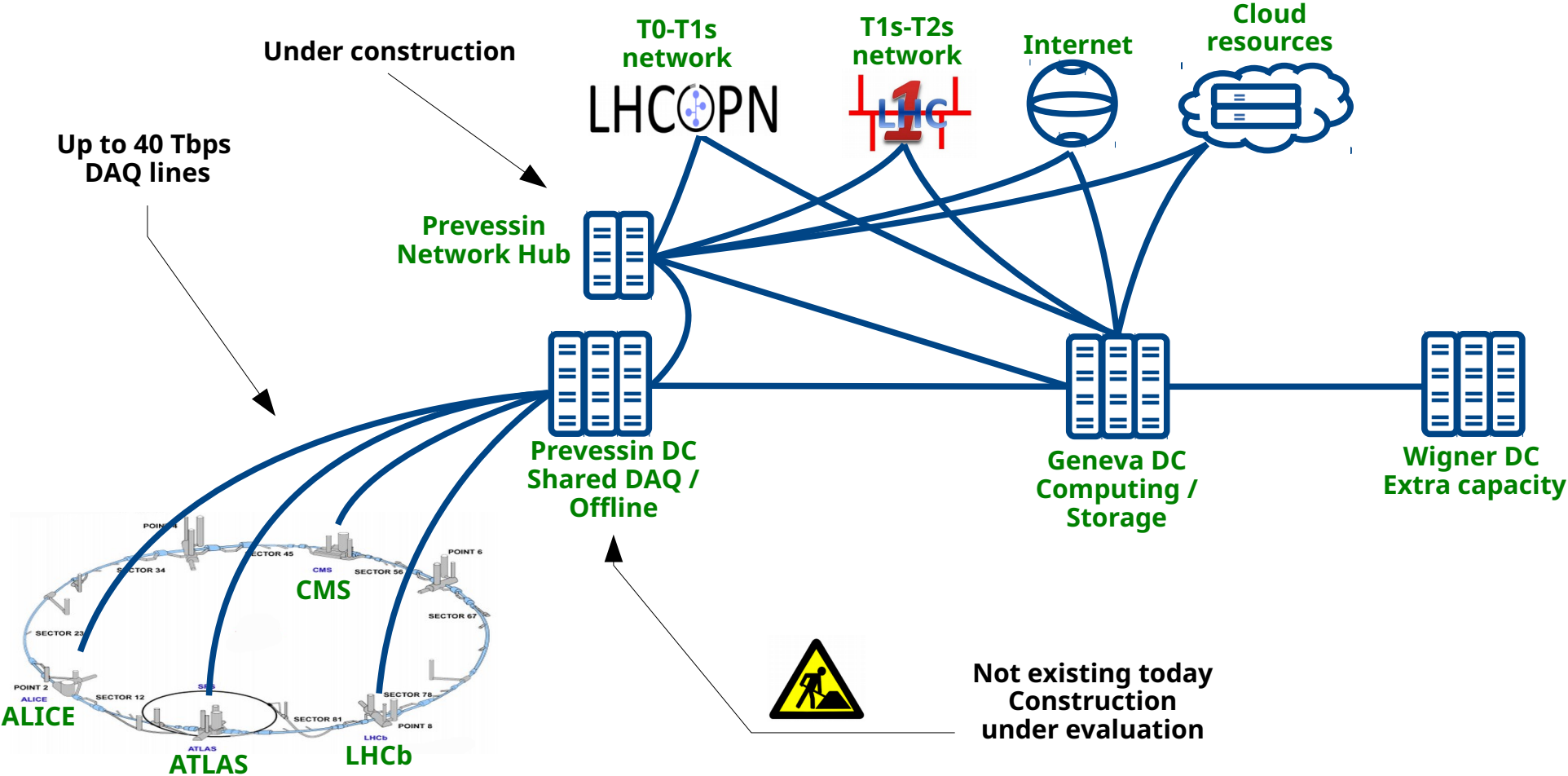
2nd network hub for business continuity in Preveessin

- Construction works almost completed
- Installation of network devices will start in the summer

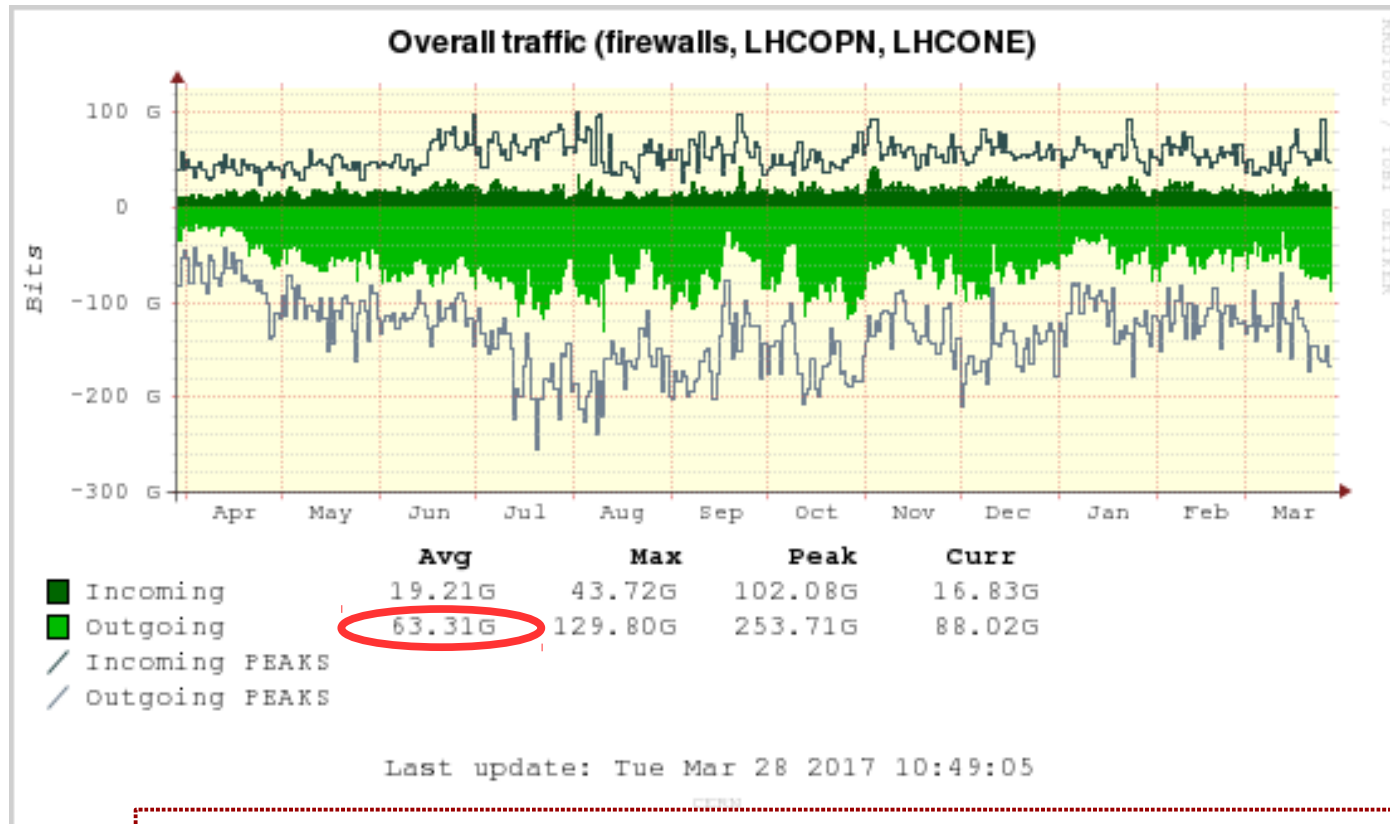
New datacentre in Preveessin

- evaluating feasibility and costs of a shared DAQ facility located in Preveessin (FR)

Possible network topology

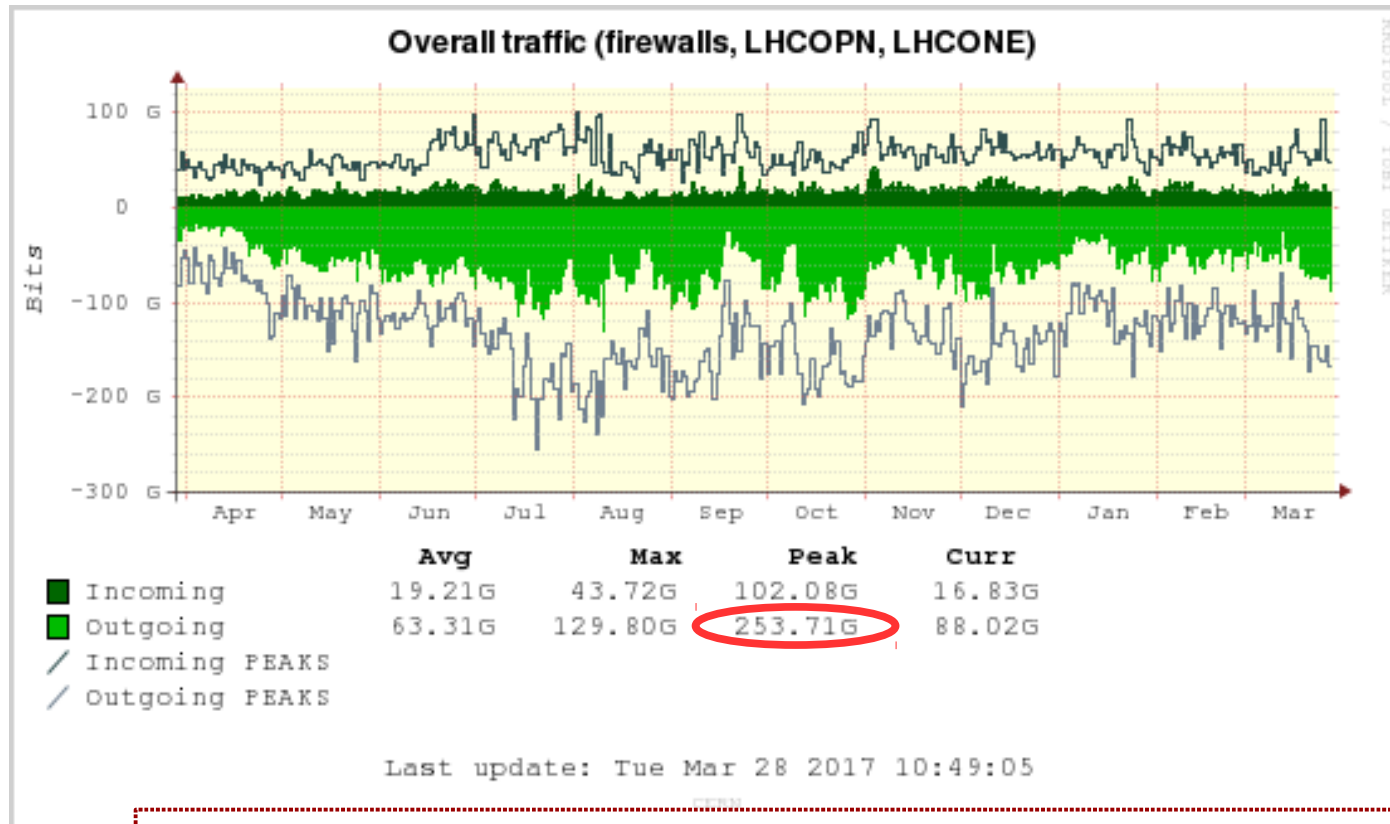


CERN external traffic – last 12 months



Average bandwidth increased of ~80% compared to previous year

CERN external traffic – last 12 months



Peak of 250Gbps outbound

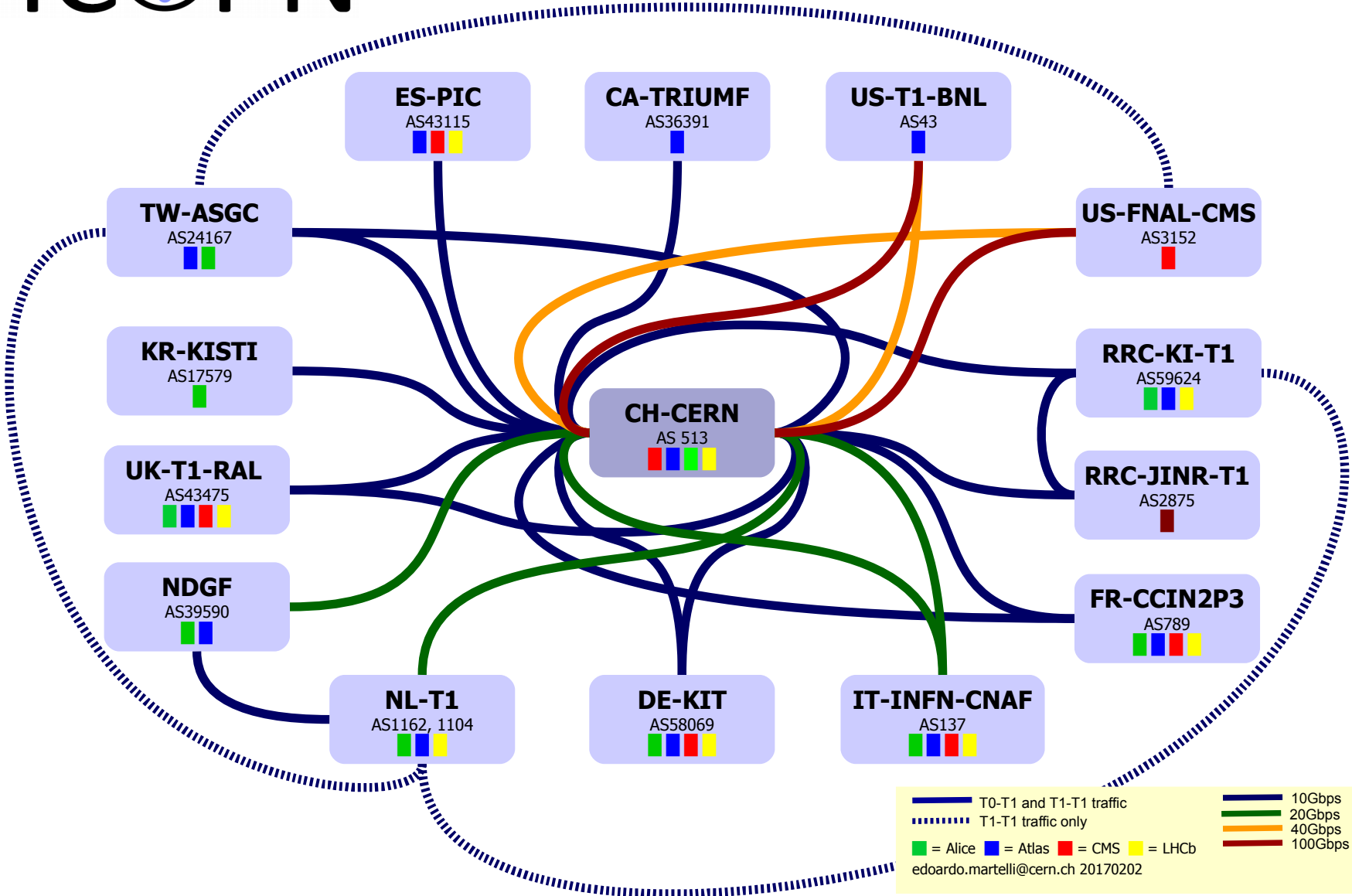
LHCOPN

Private network connecting Tier0 and Tier1s

- Reserved to LHC data transfers and analysis
- Single and bundled long distance 10G and 100G Ethernet links
- Star topology
- BGP routing: communities for traffic engineering, load balancing
- Security: only declared IP prefixes can exchange traffic.

<https://twiki.cern.ch/twiki/bin/view/LHCOPN/WebHome>

LHCOPN



Latest developments

IPv6 adoption:

- 10 Tier1s and the Tier0 now have IPv6 connectivity
- dual-stack perfSONAR installed in all of them

Doubled link capacity during last year:

- NL-T1 (2x10G)
- FR-IN2P3 (2x10G)
- NDGF (2x10G)
- IT-INFN-GARR (4x10G)
- UK-T1-RAL (2x10G, load balancing on existing backup link, now procuring 3rd 10G link)

CA-TRIUMF: is moving to SFU; TRIUMF dismantled by 2020

Plans

Complete IPv6 deployment:

- connect all Tier1s
- make all LHCOPN perfSONAR probes dual-stack
- use IPv6 for production data transfers

Upgrade Tier0-Tier1 links when necessary

- add additional 10G or move to 100G when cost effective

LHCONE

LHCONE L3VPN: Routed Virtual Private Network

- Worldwide private network connecting WLCG and other HEP sites at high bandwidth
- Bandwidth dedicated to HEP data transfers, no contention with other research projects
- Trusted traffic that can bypass slow perimeter firewalls

<https://twiki.cern.ch/twiki/bin/view/LHCONE/WebHome>



LHCONE L3VPN - status

- Over 20 national and international Research Networks
- Interconnections at Open Exchange Points including NetherLight, StarLight, MANLAN, WIX, CERNlight, Hong Kong and others
- Trans-Atlantic connectivity provided by ESnet, GEANT, Internet2, NORDUnet and SURFnet
- Trans-Pacific connectivity provided by ASGCnet, KREOnet, SINET
- Euro-Asia connectivity provided by GEANT, TEIN, TIFR
- ~74 end sites connected:
 - 14 Tier0/1s
 - 60 Tier2s



LHCONE L3VPN - update

Traffic within LHCONE is steadily growing

- GÉANT has seen peaks of over 100Gbps
- Growth of over 65% from Q2 2015 to Q2 2016
- ESnet sees more LHCONE traffic than LHCOPN

Some NRENs and sites need to upgrade their connection

- GÉANT is already working with the NRENs for the upgrades

Expected to see further increases after the upgrades



LHCONE L3VPN - update

The LHCONE network is expanding

- Ukrain, Russia and Belgium connected to GÉANT (Russian ban finally lifted)
- Brazil connected to GÉANT, Chile is interested to join
- TEIN (Asia) connected to GÉANT, provides transit to Thailand



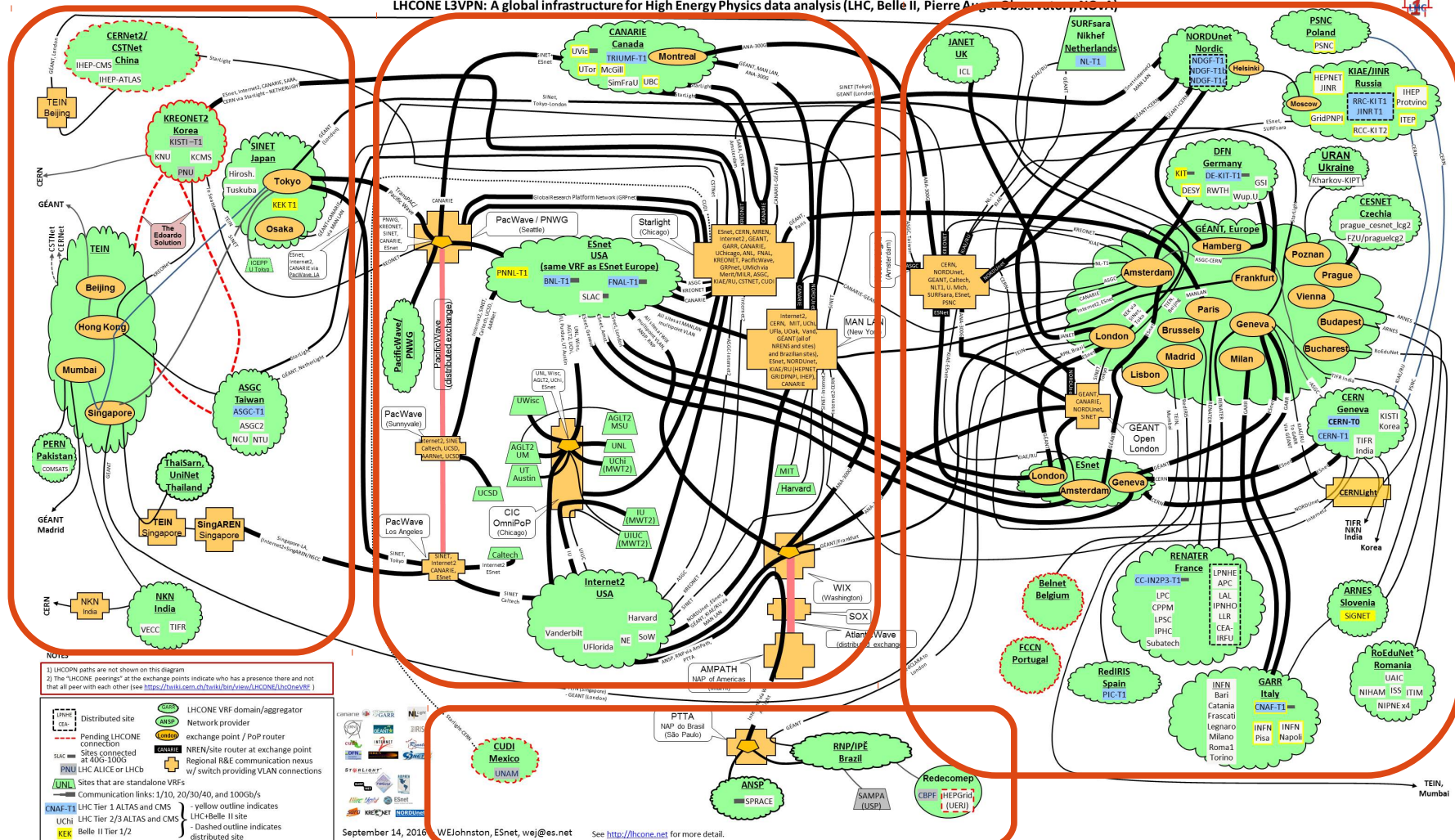
LHCONE L3VPN - current topology

Asia

North America

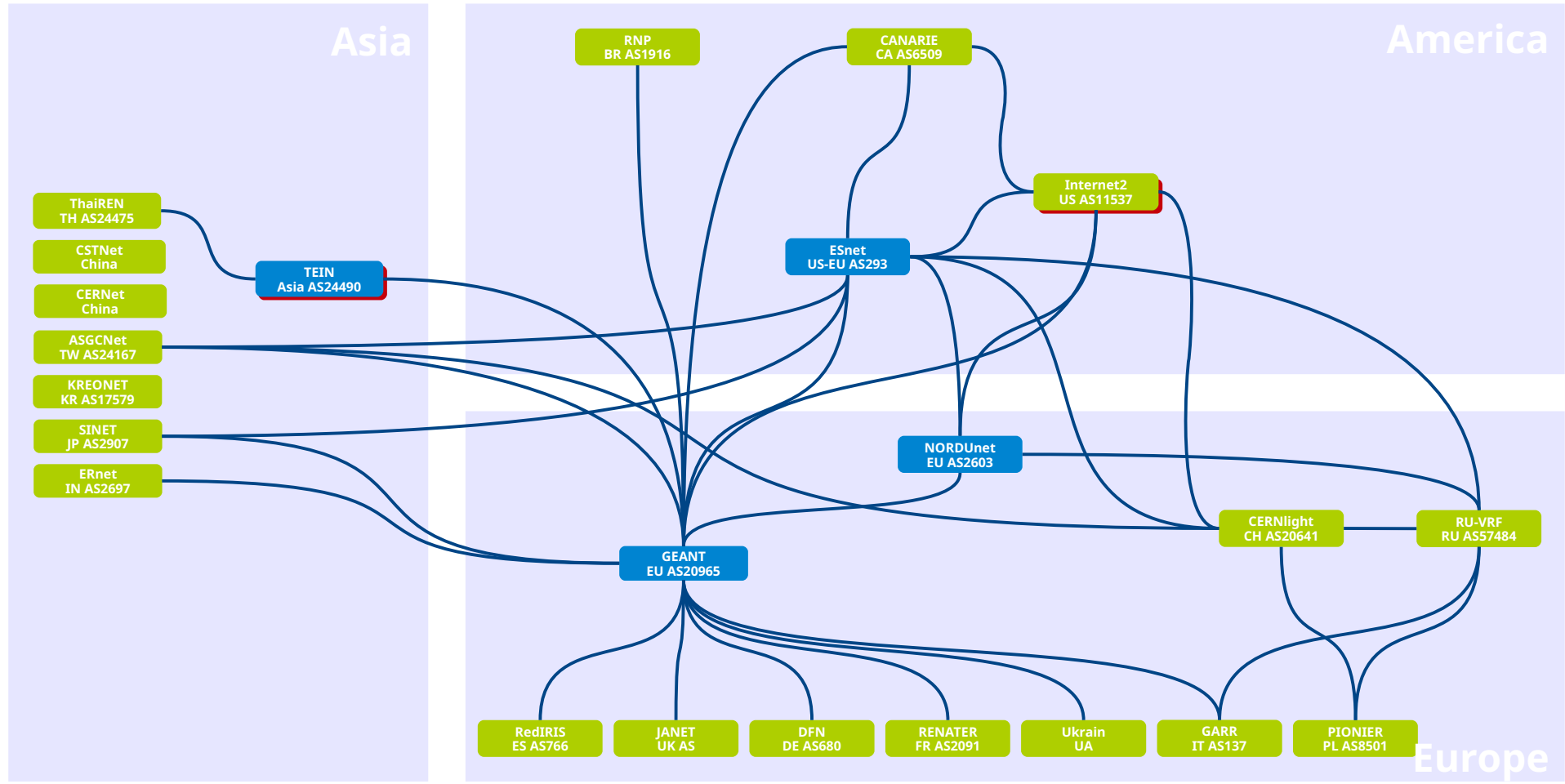
Europe

LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger, OPERA, etc.)



South America

Map of VRFs



International VRF

National VRF

Transit

LHCONE development in Asia



SINET Japan ready to peer with TEIN in Hong Kong, but existing interconnection turned out to be just 1Gbps. Will wait till TEIN's routers are upgrade in June 2017



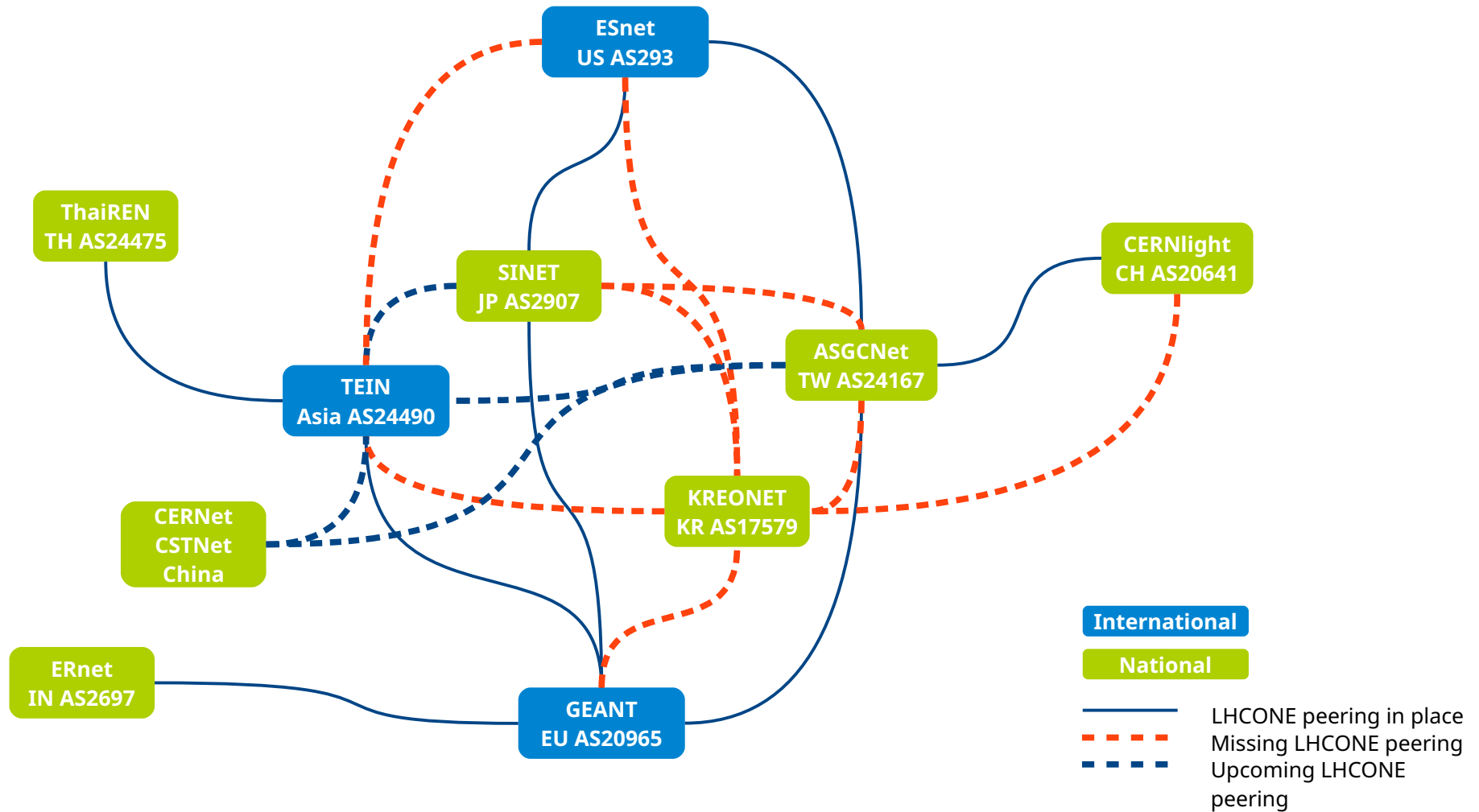
Other peerings with TEIN are waiting for the TEIN's routers upgrade



IHEP and CCNU interested to connect to LHCONE. Discussion started with CERnet and CSTnet. CSTnet has promised to implement LHCONE



Upcoming developments



TransPAC

Cooperative partnership among Indiana University, APAN, TEIN, JGN/NICT-Japan, NII-Japan, CERNET, and others.

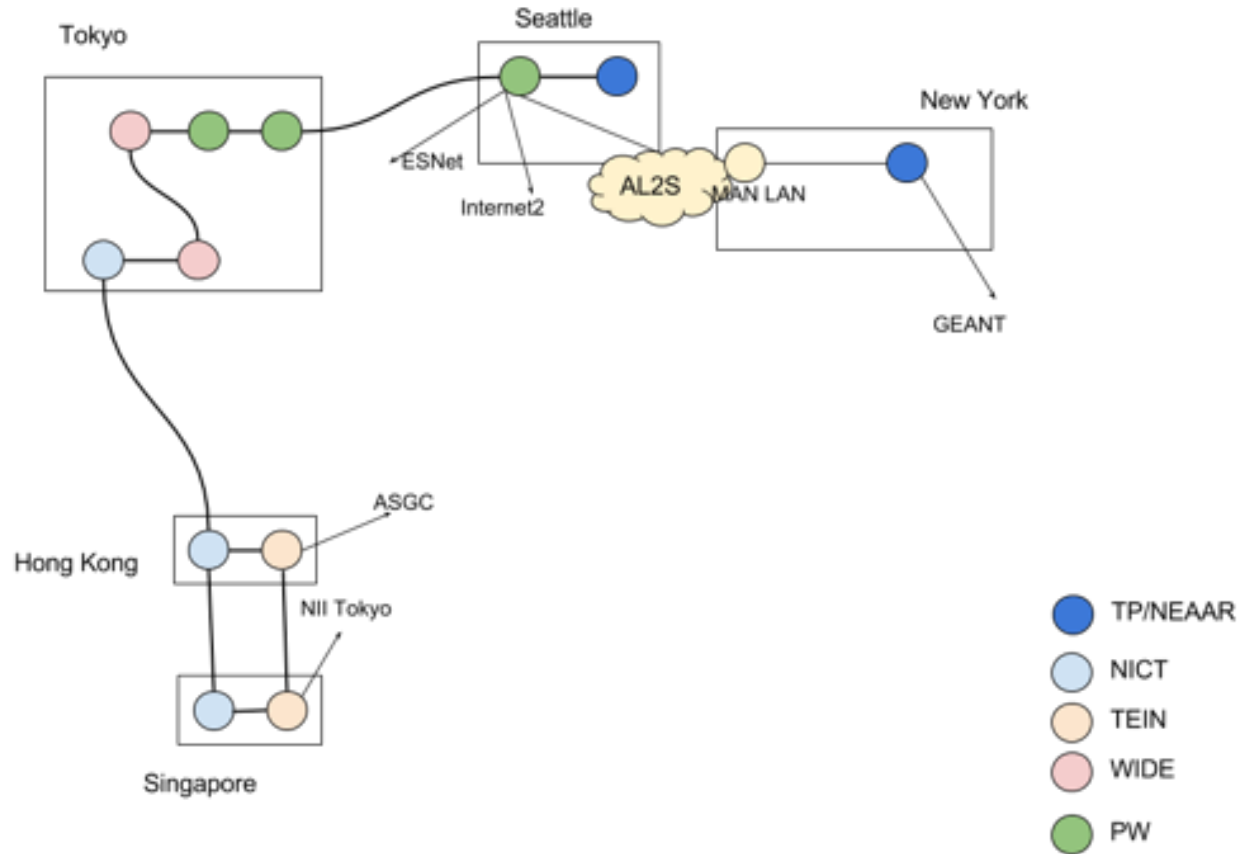
Started in March 2015, running through Feb 2020

\$4.8M over 5 years. Includes funding for circuits (100G), exchange points, application support/engagement, research

Increased engagement with the wider community through groups such as APAN, TEIN/Asi@Connect, GNA, GLIF, LHCONE, Internet2, etc

Offers capacity over transpacific 100G link to LHCONE, to connect TEIN with Internet2 and ESnet

LHCONE connectivity via TransPAC



ALICE sites connected to LHCONE

ALICE sites:

- ~ 75 sites around the world
- at least 35 sites connected to LHCONE



ALICE

ALICE sites



Site	Group	Tier	LHCONE
Vienna	Austria	T2	n
CBPF	Brasil	T2	y
SaoPaulo	Brasil	T2	
CERN (Meyrin)	CERN	T0	y
CERN (Wigner)	CERN	T0	y
Prague	Czech Republic	T2	y
FZK	Germany	T1	y
GSI	Germany	T2	y
Athens	Greece	T2	n
KFKI	Hungary	T2	n
CCIN2P3	IN2P3	T1	y
Clermont	IN2P3	T2	y
Grenoble	IN2P3	T2	y
GRIF_IPNO	IN2P3	T2	y
GRIF_IRFU	IN2P3	T2	y
IPNL	IN2P3	T2	
Strasbourg_IRES	IN2P3	T2	y
Subatech_CCIPL	IN2P3	T2	y
Kolkata	India	T2	y
Bandung	Indonesia	T2	n
Cibinong	Indonesia	T2	n
Bari	INFN	T2	y
Cagliari	INFN	T2	
Catania	INFN	T2	y
CNAF	INFN	T1	y

ALICE sites



Site	Group	Tier	LHCONE
Legnaro	INFN	T2	y
Torino	INFN	T2	y
Trieste	INFN	T2	
TriGrid_Catania	INFN	T2	y
Hiroshima	Japan	T2	n
Tsukuba	Japan	T2	n
UNAM	Mexico	T2	n
CSC	Nordic Countries	T2	
DCSC_KU	Nordic Countries	T2	
HIP	Nordic Countries	T2	
LUNARC	Nordic Countries	T2	
SNIC	Nordic Countries	T2	
UiB	Nordic Countries	T2	
COMSATS	Pakistan	T2	n
PAKGRID	Pakistan	T2	n
Cyfronet	Poland	T2	y
Poznan	Poland	T2	
WUT	Poland	T2	y
IHEP	RDIG	T2	y
ITEP	RDIG	T2	y
JINR	RDIG	T2	y
MEPHI	RDIG	T2	
PNPI	RDIG	T2	y
RRC-KI	RDIG	T2	y
SPbSU	RDIG	T2	

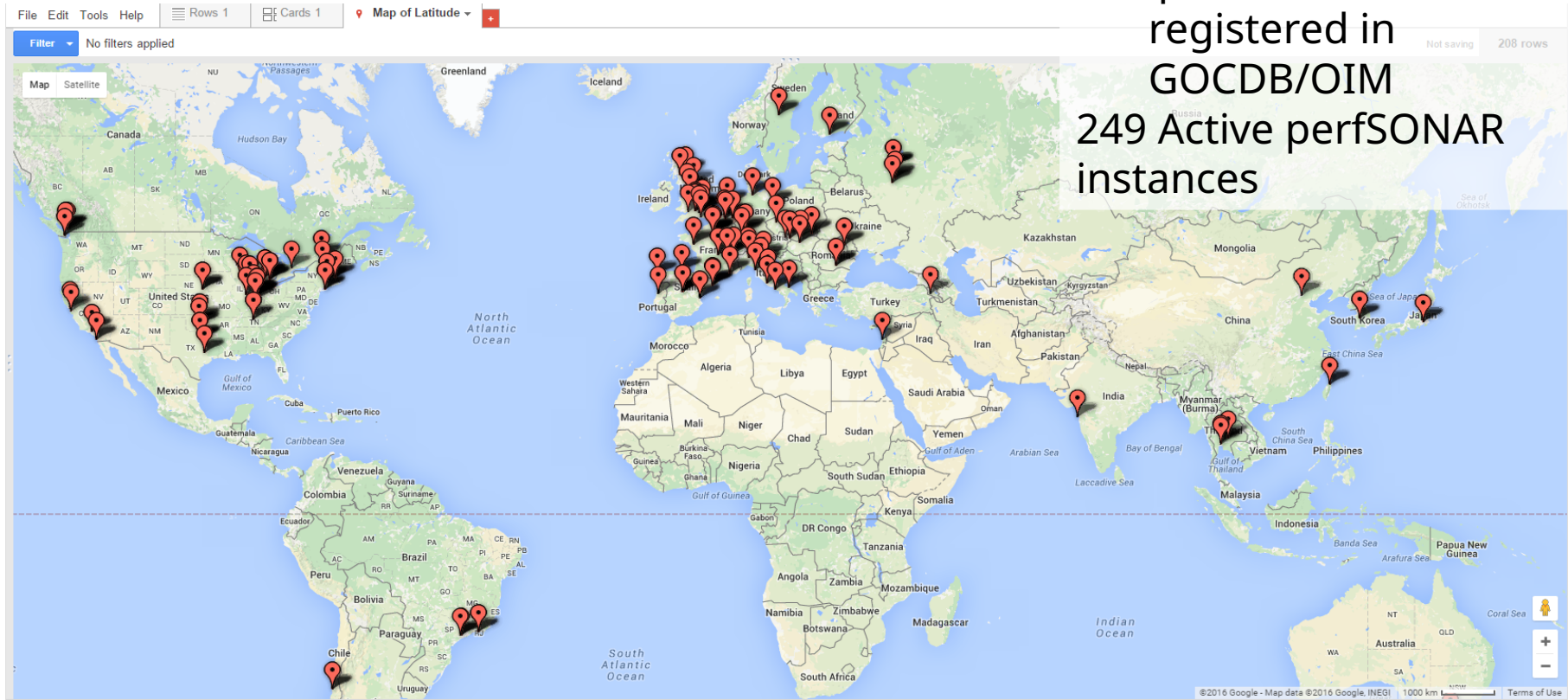
ALICE sites



Site	Group	Tier	LHCONE
Troitsk	RDIG	T2	y
KISTI	Republic of Korea	T2	n
ISS	Romania	T2	
NIHAM	Romania	T2	y
NIPNE	Romania	T2	y
UPB	Romania	T2	
Bratislava	Slovakia	T2	
Kosice	Slovakia	T2	
ZA_CHPC	South Africa	T2	
Madrid	Spain	T2	y
Trujillo	Spain	T2	
NECTEC	Thailand	T2	
SUT	Thailand	T2	y
NIKHEF	The Netherlands	T2	y
SARA	The Netherlands	T1	y
Birmingham	UK	T2	n
Oxford	UK	T2	n
RAL	UK	T1	n
BITP	Ukraine	T2	
ICYB	Ukraine	T2	
ISMA	Ukraine	T2	
KNU	Ukraine	T2	
LBL	US	T2	
LLNL	US	T2	
ORNL	US	T2	

WLCG perfSONAR

278 perfSONAR instances
registered in
GOCDB/OIM
249 Active perfSONAR
instances



perfSONAR - update

Mesh have been reconfigured/optimized. There are now experiment specific meshes: ATLAS, CMS and LHCb which allow daily bandwidth tests. ALICE uses MonaLISA

All the LHCOPN and LHCONE data is available in ElasticSearch via OSG (ongoing)

Some degradation in the data retrieving of the LHCOPN/ONE MaDDash. Upcoming upgrade to v4 should be the opportunity to make the nodes more resilient

perfSONAR - update

ATLAS is working on getting network metrics into an analytics platform:

- packet loss from perfSONAR
- network stats and flows from ESnet routers
- stats from CERN routers will be added

Looking for more router sources

Completed MCA (Mesh Configuration Admin) tool. Could be used by WLCG

WLCG is working on ETF (Experiment Test Framework) to monitor perfSONAR services

Open to other HEP collaborations

The L3VPN is now used also by:

- **Belle II experiment**



- **NOvA neutrino experiment**



- **Pierre Auger observatory**



- **XENON dark matter project**



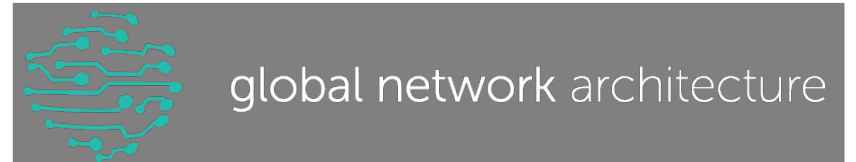
GNA - Global Network Architecture

Sharing of R&E network resources for common benefit

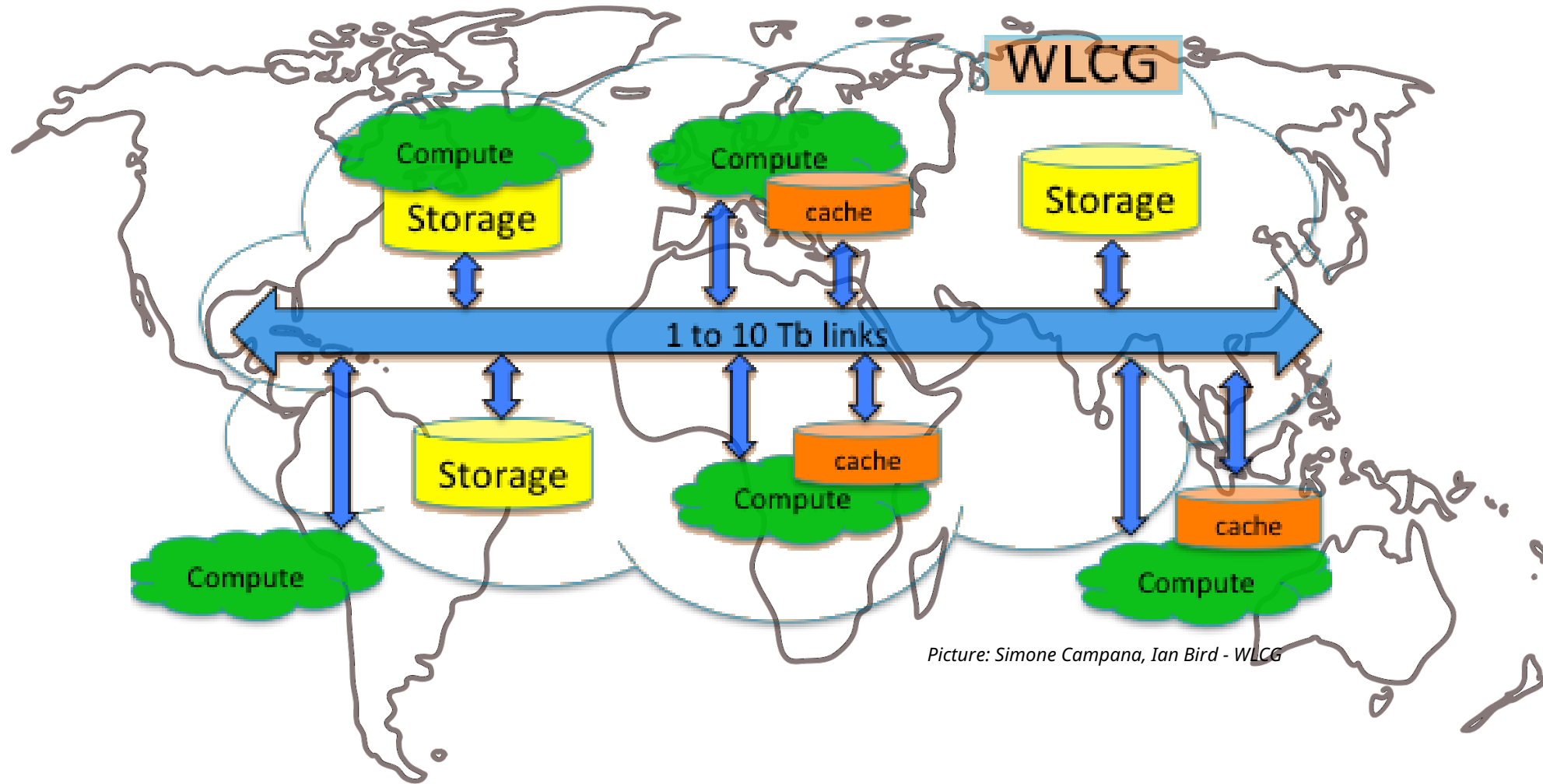
Extend benefits gained with ANA-300G experience
(sharing of three 100G transatlantic links)

Pillars:

- resource sharing
- aligning operational standards
- maximizing funding and investments
- knowledge sharing and outreach
- increasing global collaboration



Possible change of computing model

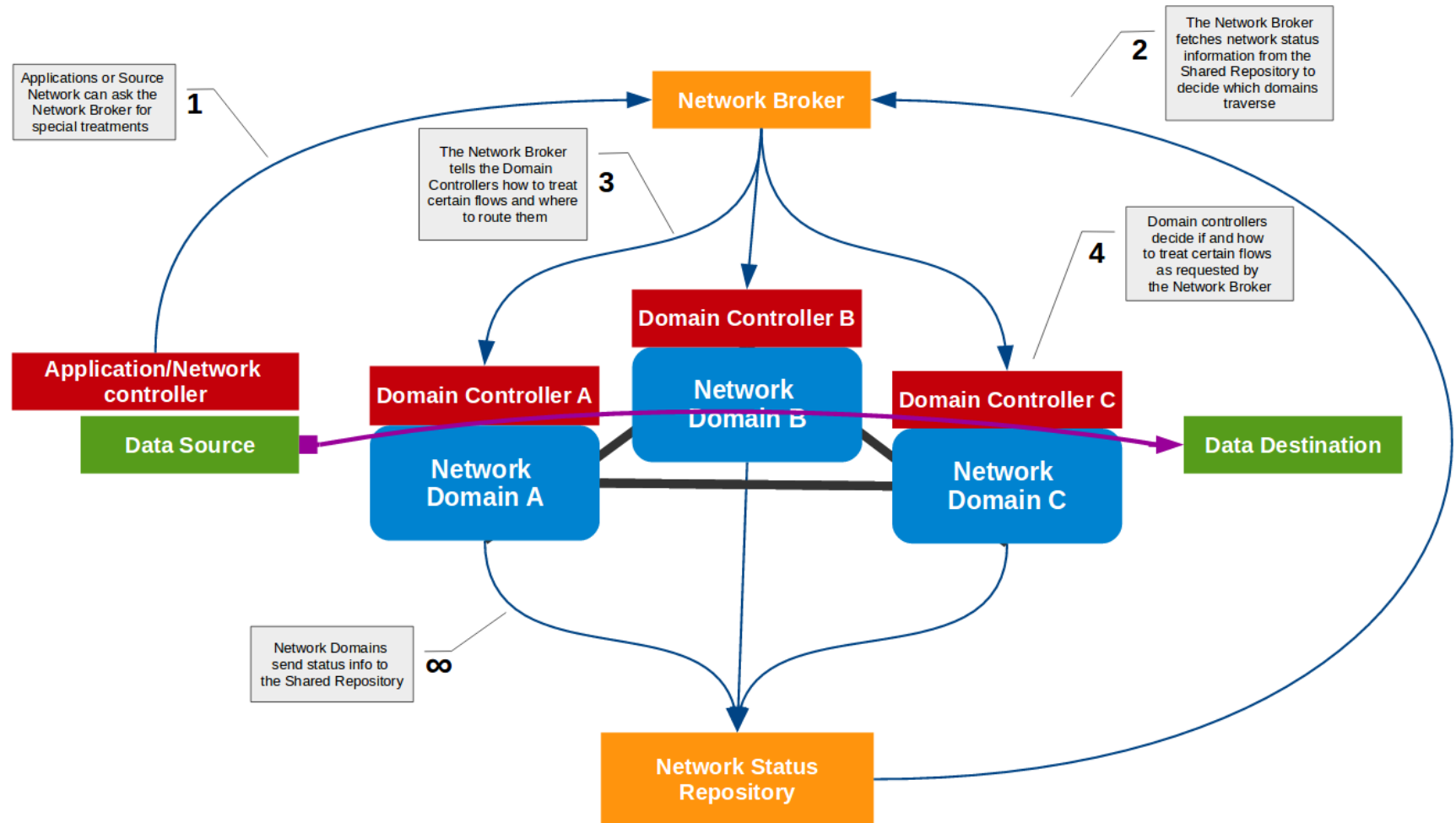


Picture: Simone Campana, Ian Bird - WLCG

"Consolidate storage in few locations to decrease costs"

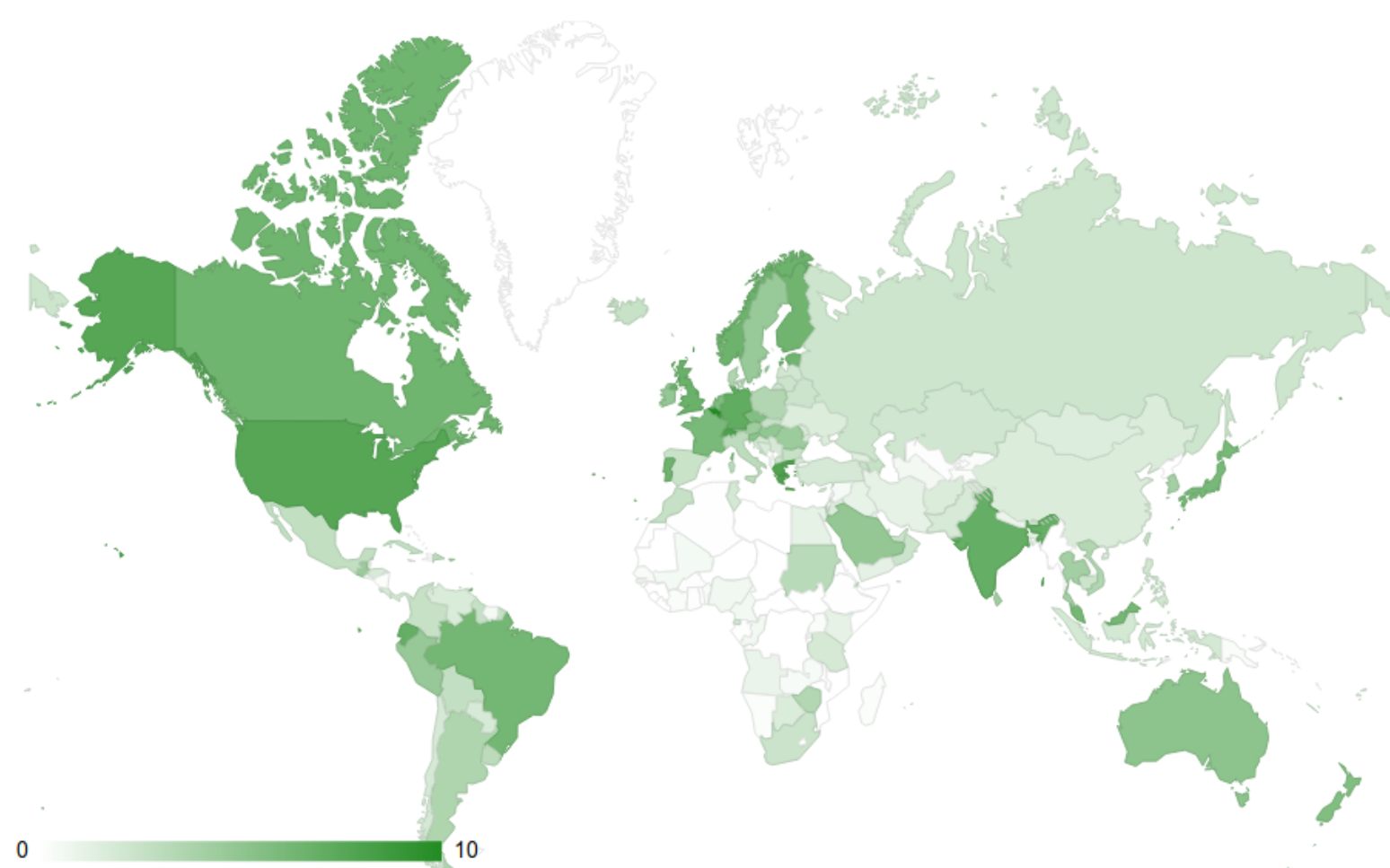
Networking for Run4

Proposed project to improve network utilization efficiency



IPv6 deployment

IPv6 adoption



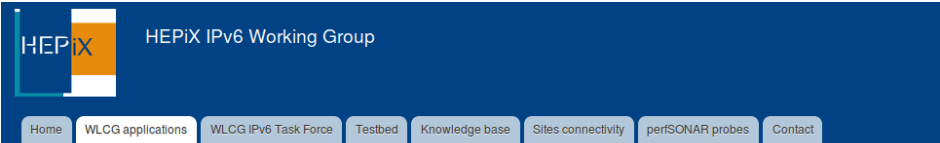
<http://6lab.cisco.com/stats/>

HEPiX IPv6 working group

IPv6 deployment have been followed up by the HEPiX IPv6 working group

It surveys IPv6 compliance of HEP applications and IPv6 conenctivity of HEP sites

Lists are maintained at the web site: <http://hepixonweb.cern.ch/>



Home

IPv6 compliance of WLCG applications

Software Component	Type	Used by Experiment	Version	IPv6 Compliance
ALIEN	LHC Experiment Application	ALICE		YES
ARC CE	Middleware	ATLAS, CMS		Unknown
ARGUS	Middleware	ALICE, ATLAS, CMS, LHCb		Unknown
BDII	Middleware	ATLAS, CMS, LHCb	EMI 2	YES
BestMAN	Middleware	ATLAS, CMS		YES
CASTOR	Middleware	ALICE, ATLAS, CMS, LHCb		NO
clengine	Monitoring			Unknown
CMS Tag Collector	LHC Experiment Application	CMS		Unknown
CMSSW	LHC Experiment Application	CMS		Unknown
cmsweb	LHC Experiment Application	CMS		YES
CRAB 2	LHC Experiment Application	CMS		Unknown
Cream CE	Middleware	ALICE, ATLAS, CMS, LHCb	1.16.2	YES



Home

Sites IPv6 connectivity

Name	Type	LHCOPN IPv6 peering	LHCONE IPv6 peering	LHCONE IPv6 peers	NREN IPv6 peers	IPv6 LAN	dualstack perSONAR	dualstack storage percentage by 1st April 2017	dualstack storage percentage by 31st May 2017	dualstack storage percentage by 31st July 2017	Network Statistics
UKI-LT2-IC-HEP	Tier2		Yes	Janet	Janet	Yes		100			
TRIUMF	Tier1	Yes	Yes	Canarie	BCNET	Yes	Yes	0			
RAL	Tier1	No			JANET	No	Yes	0			
pragueicg2 (FZU)	Tier2	No	Yes	CESNET	CESNET	Yes	Yes	100			http://netreport.cesnet.cz/netreport/hep-cesnet-experimental-facility2/
NL-T1-Nikhef	Tier1	Yes	No	GEANT	SURFnet	Yes	No				



IPv6 only worker nodes

IPv6 requirements for LHC Experiments have been discussed at the GDB

A plan to support IPv6 only worker nodes has been proposed.

The aims are:

- provide a migration path for sites
- allow long term planning for network setups
- give the possibility to make use of IPv6 only resources, should they become available

It is essential to provide dual-stack access to storage



Dual-stack storage

Most storage solutions and protocols now work in dual-stack mode

- dCache, DPM, StoRM, FTS
- XrootD 4, GridFTP, WebDAV

Several sites have been running dual-stack for some time

Production data transfers over IPv6 have started happening

IPv6 at Tier0-Tier1s

- Good IPv6 adoption
- 10 Tier1s and the Tier0 peering over IPv6
- dual-stack perfSONAR installed in all of them

Still missing:

- KR-KISTI (work in progress)
- RRC-KI-T1 KIAE (work in progress)
- RRC-KI-T1 JINR (will follow KIAE)

No IPv6 storage in most of them though, including CERN

IPv6 at Tier2s

IPv6 connectivity provide by most of the LHCONE VRFs
12 Tier2s have IPv6 connectivity with LHCONE

HEPiX is writing an How-To deploy IPv6 at a Tier2:
<https://hepixon.web.cern.ch/content/how-deploy-ipv6-wlcg-tier-2-site>

An IPv6 tutorial is scheduled for the WLCG workshop in Manchester in June

IPv6 and ALICE

ALICE central services have been dual stack for more than a year

Storage is fully federated based on XrootD. XrootD 4.1 is needed to support IPv6

To support IPv6-only resources, all data must be available on some IPv6-enabled storage

ALICE can support IPv6-only CPU resources as soon as enough sites have upgraded their storage to dual stack. At the moment only 5% have done it. Aiming to all sites dual-stack by the end of Run II



IPv6 Monitoring

Network performance on perfSONAR

<http://psmad.grid.iu.edu/maddash-webui/index.cgi?dashboard=Dual-Stack%20Mesh%20Config>

FTS transfers in MONIT

<https://monit.cern.ch/>

(add "data.ipv6:true" to any FTS monitor page for IPv6 only traffic)

ETF

https://etf-ipv6-dev.cern.ch/etf/check_mk/

Dual-Stack Mesh Config - IPv6 Latency Test

Loss rate is <= 0 Loss rate is >= 0 Loss rate is >= 0.01 Unable to retrieve data Check has not yet run

Found a total of 10 problems involving 10 hosts in the grid



Clouds activities

HELIX NEBULA the Science Cloud

Procurers: CERN, CNRS, DESY, EMBL-EBI, ESRF, IFAE, INFN, KIT, SURFSara, STFC

- Procurers have committed funds (>1.6M€), manpower, use-cases with applications & data, in-house IT resources

Objective: procure innovative IaaS level cloud services

- Fully and seamlessly integrating commercial cloud (IaaS) resources with in-house resources and European e-Infrastructures
- To form a hybrid cloud platform for science

Services will be made available to end-users from many research communities: High-energy physics, astronomy, life sciences, neutron/photon sciences, long tail of science

Co-funded via H2020 (Jan'16-Jun'18) as a Pre-Commercial Procurement (PCP) project: Grant Agreement 687614, total procurement volume: >5M€



HNSciCloud challenges

Innovative IaaS level cloud services integrated with procurers in-house resources and public e-infrastructure to support a range of scientific workloads

Compute and Storage

support a range of virtual machine and container configurations including HPC working with datasets in the petabyte range

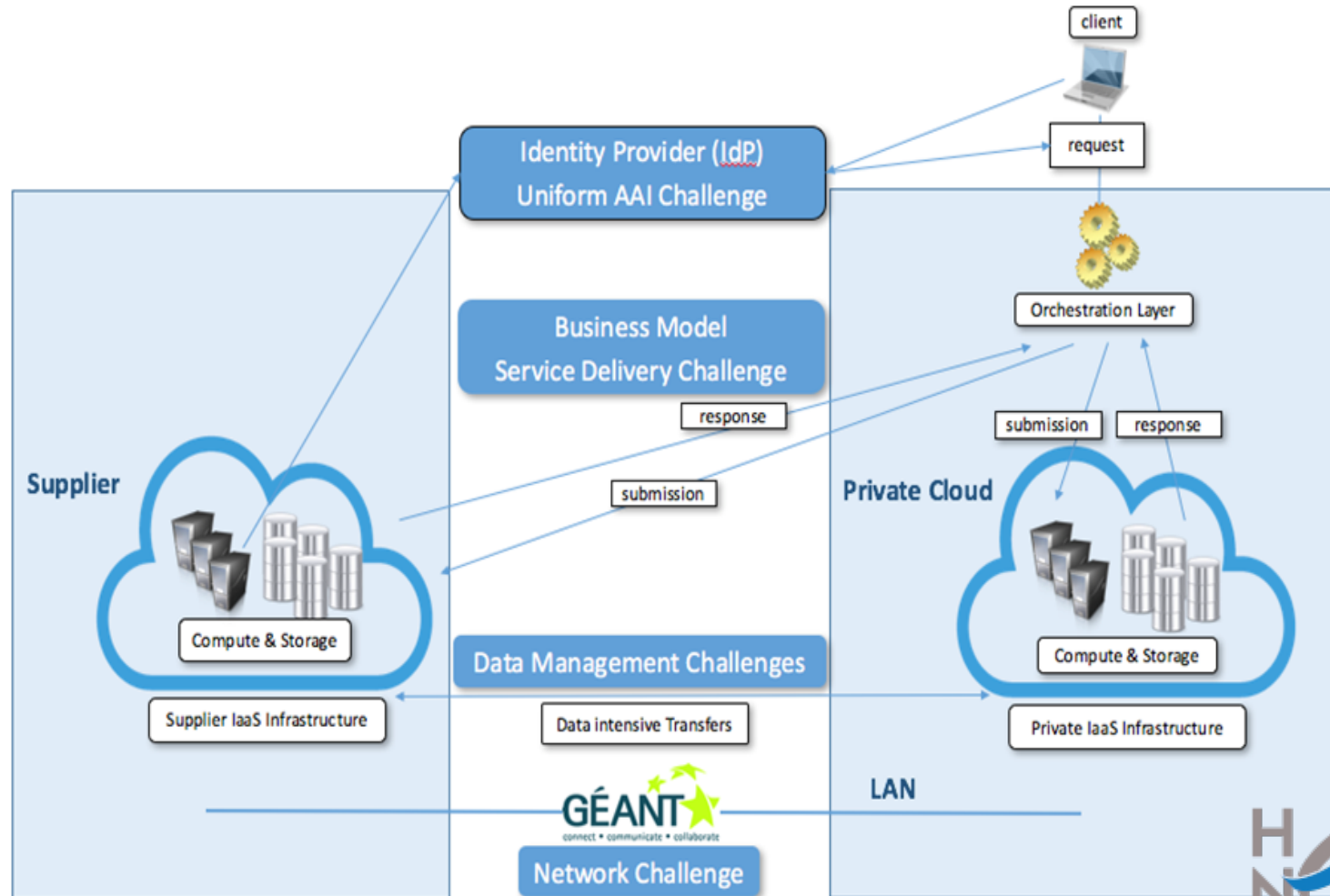
Network Connectivity and Federated Identity Management

provide high-end network capacity via GEANT for the whole platform with common identity and access management

Service Payment Models

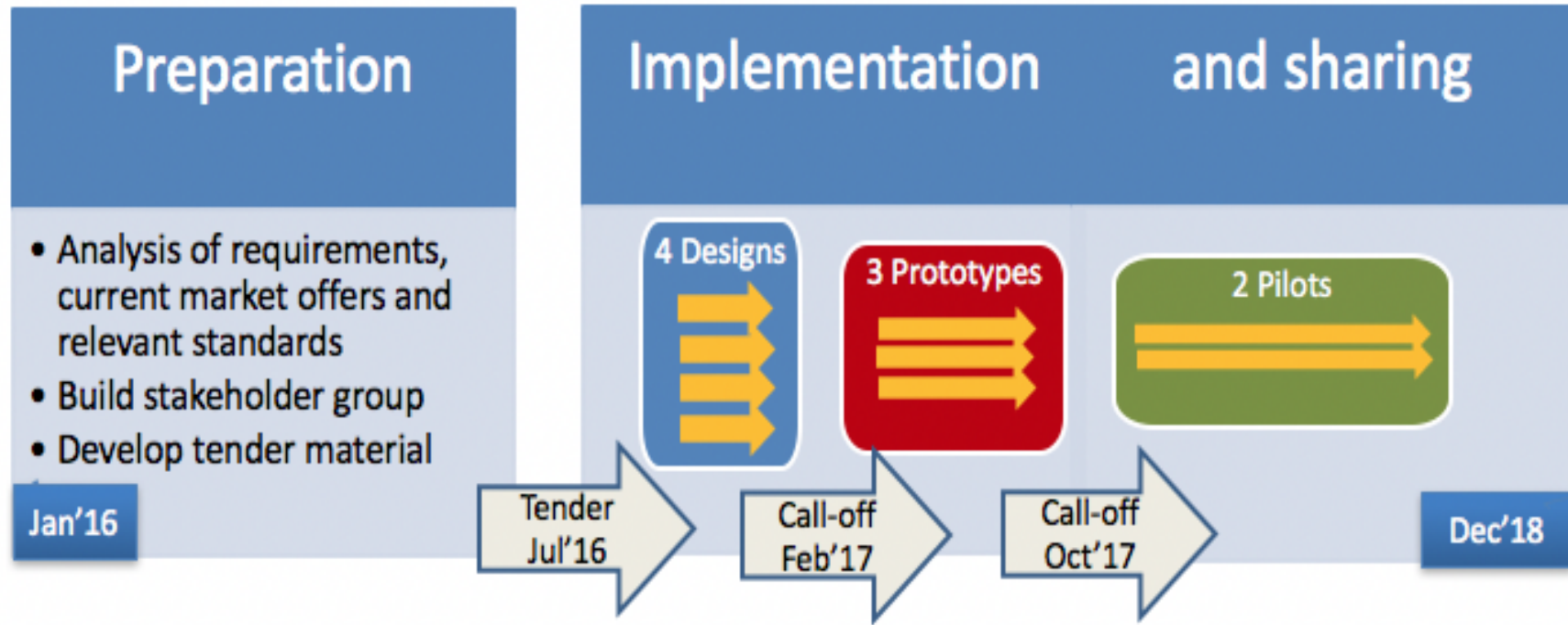
explore a range of purchasing options to determine those most appropriate for the scientific application workloads to be deployed

High Level Architecture of the Hybrid Cloud Platform including the R&D challenges



Project phases

We are here



Each step is **competitive** - only contractors that successfully complete the previous step can bid in the next

Selected bidders

Admitted to the Design Phase

- T-Systems, Huawei, Cyfronet, Divia
- IBM
- RHEA Group, T-Systems, exoscale, SixSq
- Indra, HPE, Advania, SixSq

Other major players not interested or dropped out just before tender submission

Admitted to the Prototype Phase

- T-Systems, Huawei, Cyfronet, Divia
- IBM
- RHEA Group, T-Systems, exoscale, SixSq

Current Status: Evaluation of Bids in the Prototype Phase

Prototype phase

Collect a number of tests from the Procurers

- To test/validate the design proposed for the several project challenges
- To be able to assess results of the tests as one of the main criteria to be eligible to the next phase (Pilot)

Tests consist in applications and some selected workloads

Next phase (Pilot) will be based on the Prototype to be assessed on 3 main aspects:

- Performance
- Scalability
- Security

CMS GoogleCloud engine demo at SC 2017

Challenge: Doubling CMS computing using Google Cloud Engine
Expand the Fermilab facility to an additional 160,000 cores
Use HEPCloud technology to do this as transparently as possible
to the application

Live demo during Supercomputing 2016
730172 simulation jobs submitted; only 47 did not complete
205 M physics events generated, yielding 81.8 TB of data
Cost: ~1.6 cents per core-hour (on-premises: 0.9 cents per core-
hour assuming 100% utilization)



Google Cloud Platform

Conclusion

Summary

CERN: planning upgrades for RUN3 and for better resiliency

LHCOPN: stable, almost fully dual-stack, upgraded when needed

LHCONE: expanding, works on-going in Asia

IPv6: good progress in WAN networks, less in campus and datacentres. More dual-stack storage needed in WLCG

Clouds: testing integration with orchestration tools to fit peaks of computing demand

Questions?

edoardo.martelli@cern.ch

