



Networks for High Energy Physics: LHCOPN and LHCONE

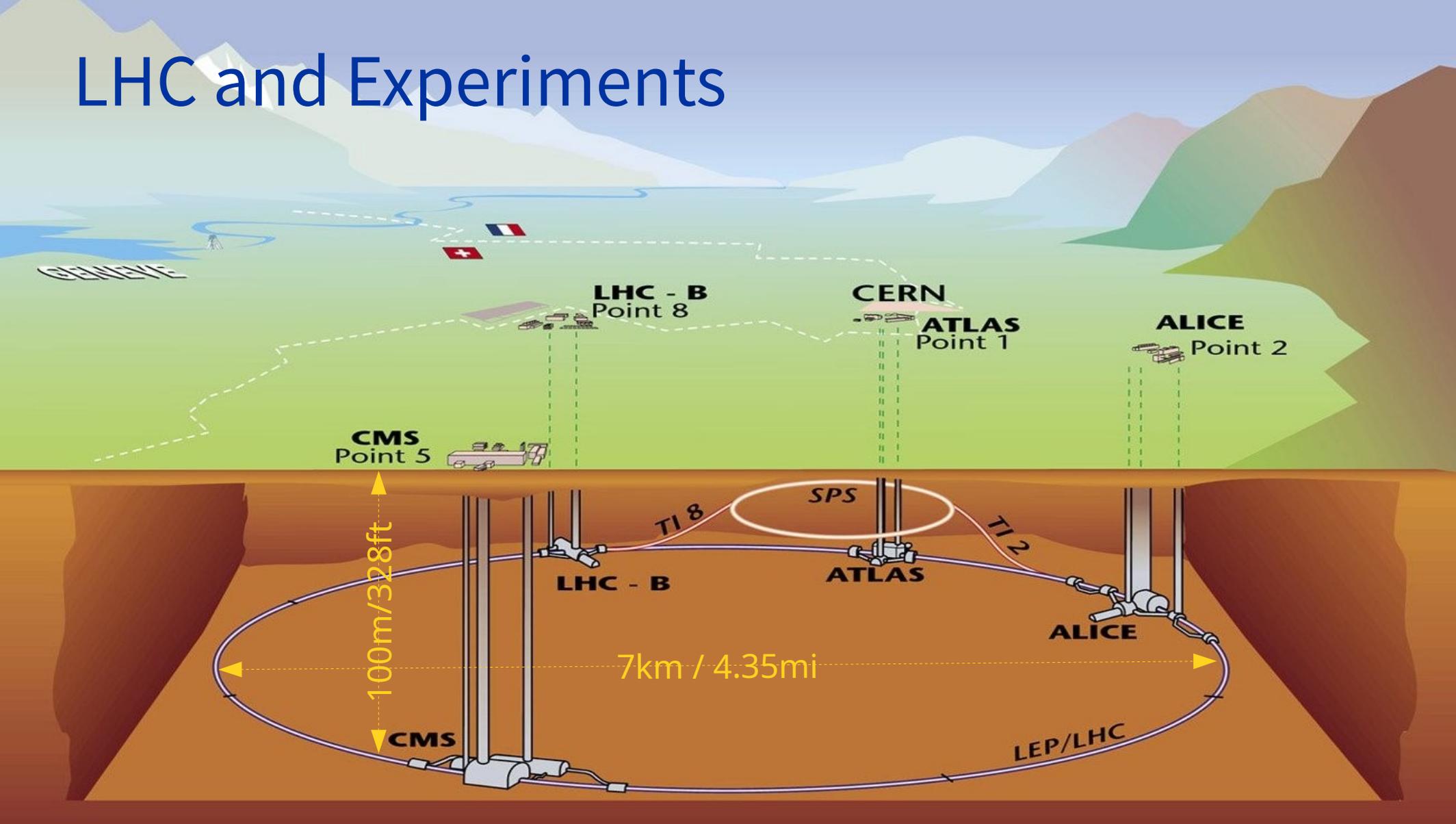
AmRP Working Group Meeting - 11 October 2022

edoardo.martelli@cern.ch

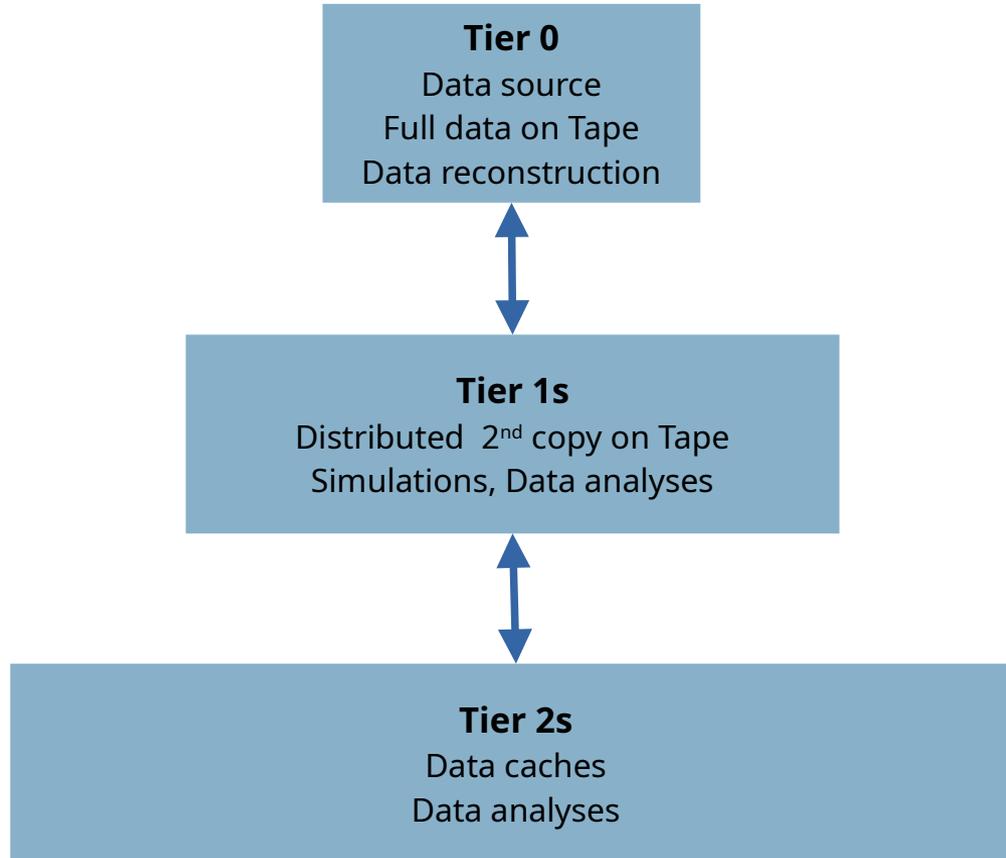


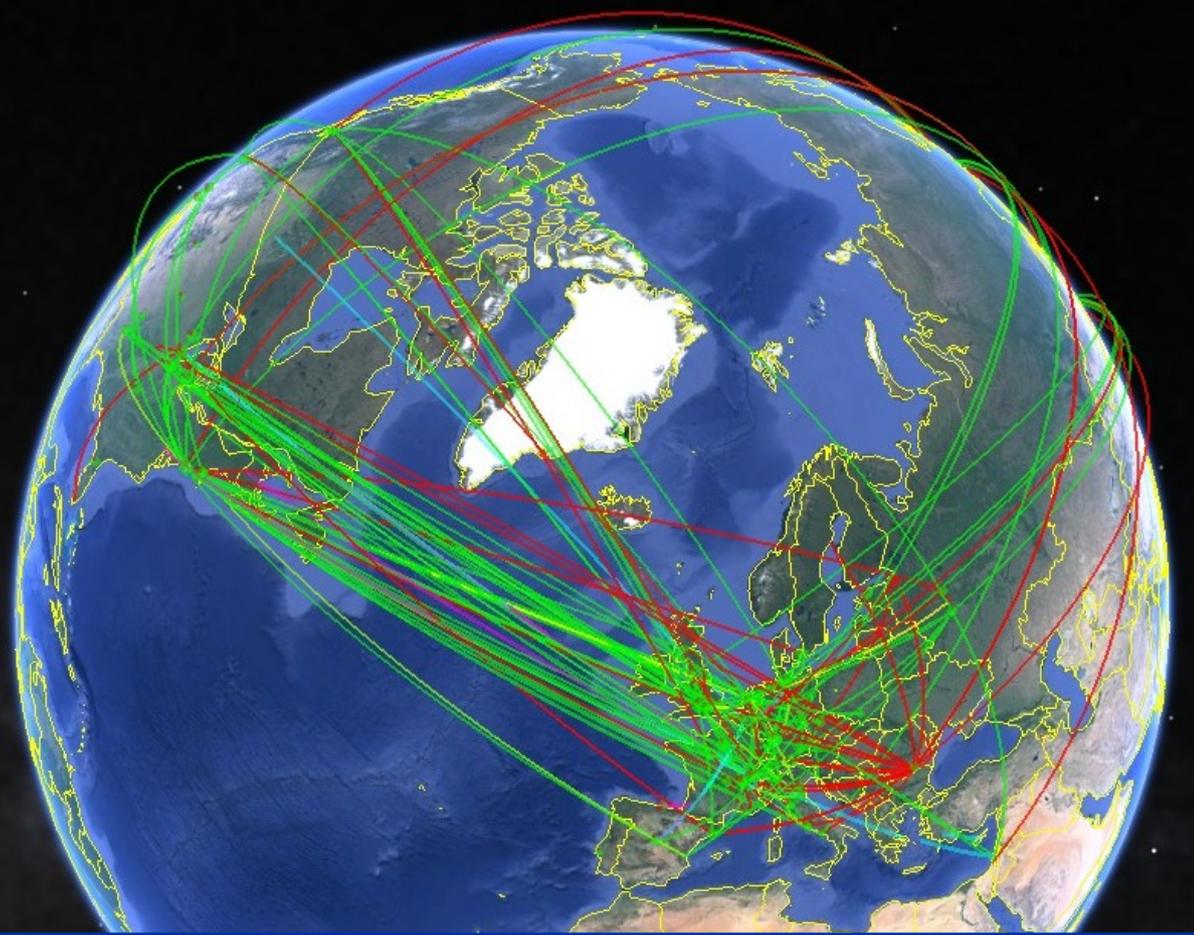
The LHC Data Challenge

LHC and Experiments



Computing Model





WLCG Networks

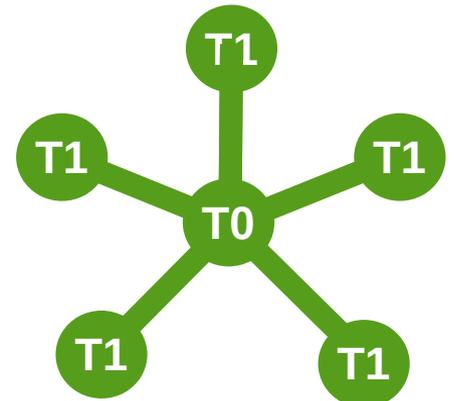
LHCOPN

LHC Optical Private Network

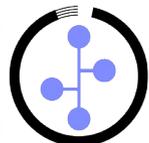
LHCOPN

Private network connecting Tier0 to all the Tier1s

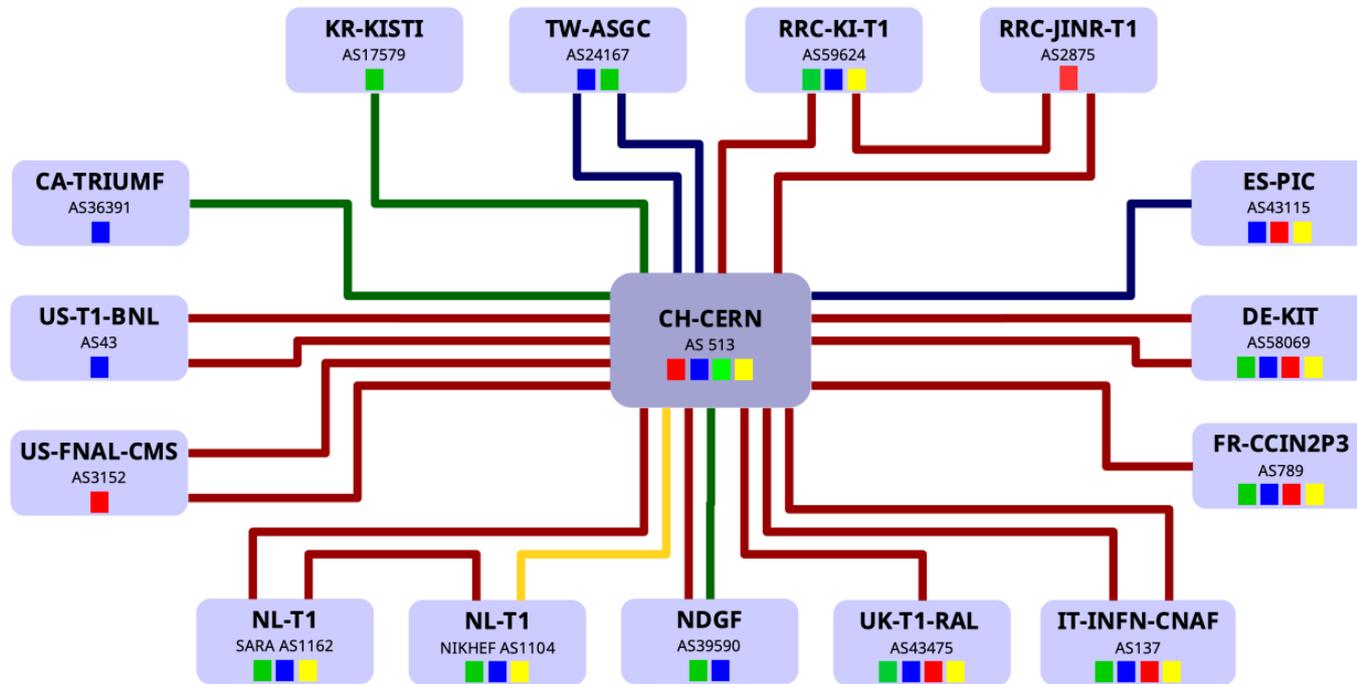
- Dedicated to LHC data transfers and analysis
- Star topology: all Tier1s connect to the Tier0
- Single and bundled long distance 10/100/400Gbps links
- Secured: only declared IP prefixes can exchange traffic
- BGP routing: communities for traffic engineering, load balancing
- Dual stack IPv4 and IPv6
- Tier1-Tier1 transit via Tier-0



LHCOPN



LHCOPN Topology

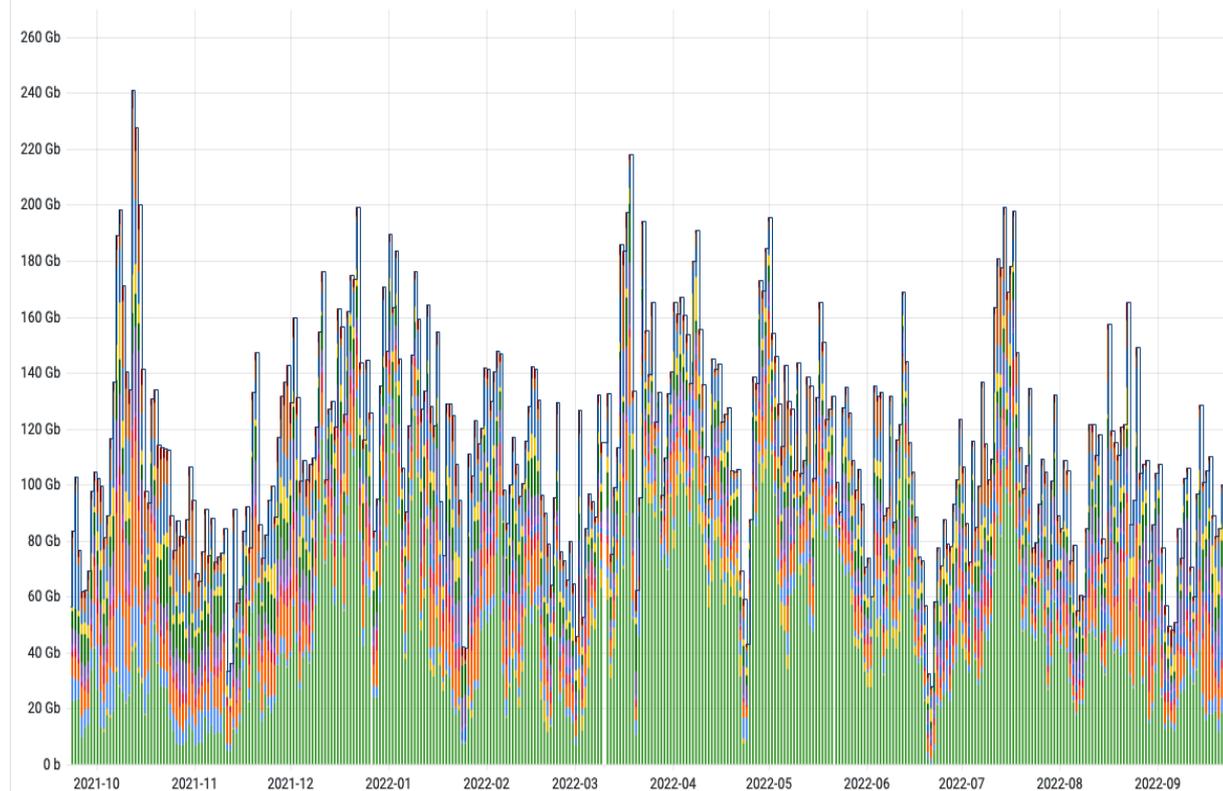


Numbers

- 14 Tier1s + 1 Tier0
- 12 countries in 3 continents
- Dual stack IPv4 IPv6
- 1.8 Tbps to the Tier0

LHCOPN Traffic – last 12 months

LHCOPN Total Traffic (CERN -> T1s)



	Mean	Last *	Max
Outgoing DE-KIT	48.4 Gb	67.0 Gb	136 Gb
Outgoing KR-KISTI	1.55 Gb	2.26 Mb	8.48 Gb
Outgoing RU-T1	5.23 Gb	2.78 Gb	32.6 Gb
Outgoing FR-IN2P3	11.2 Gb	6.62 Gb	53.2 Gb
Outgoing NDGF	4.46 Gb	2.40 Gb	14.3 Gb
Outgoing NL-T1	2.58 Gb	6.93 Gb	22.0 Gb
Outgoing TW-ASGC	1.74 Gb	2.11 Gb	5.70 Gb
Outgoing IT-INFN-CNAF	7.86 Gb	7.92 Gb	50.5 Gb
Outgoing UK-RAL	8.15 Gb	9.68 Gb	23.0 Gb
Outgoing CA-TRIUMF	4.43 Gb	2.07 Gb	24.6 Gb
Outgoing US-BNL	10.3 Gb	9.64 Gb	41.3 Gb
Outgoing US-FNAL	5.11 Gb	1.08 Gb	41.8 Gb
Outgoing ES-PIC	2.78 Gb	747 Mb	9.00 Gb
Total	114 Gb	119 Gb	241 Gb

Numbers

Moved ~450 PB
in the last 12
months

Ref: Monit Grafana



LHCOPN: latest developments

Majority of Tier1s upgraded links to 100Gbps. Few Tier1s already at 2x 100Gbps and one with 1x 400Gbps

Requirement for HL-LHC (2029): 1Tbps to each Tier1

LHCOPN capacity will be used for WLCG data challenges in preparation for Run4

LHCONE

LHC Open Network Environment

LHCONE

Private network connecting Tier1s and Tier2s:

- Serving any LHC sites according to their needs and allowing them to grow
- Model: sharing the use and the cost of expensive network resources
- A collaborative effort among Research & Education Network Providers
- Open to other HEP collaborations



LHCONE services and activities

L3VPN (VRF): routed Virtual Private Network connecting Science-DMZs - operational

Monitoring: monitoring infrastructure - operational

R&D: improve transfer speed and efficiency of network utilization

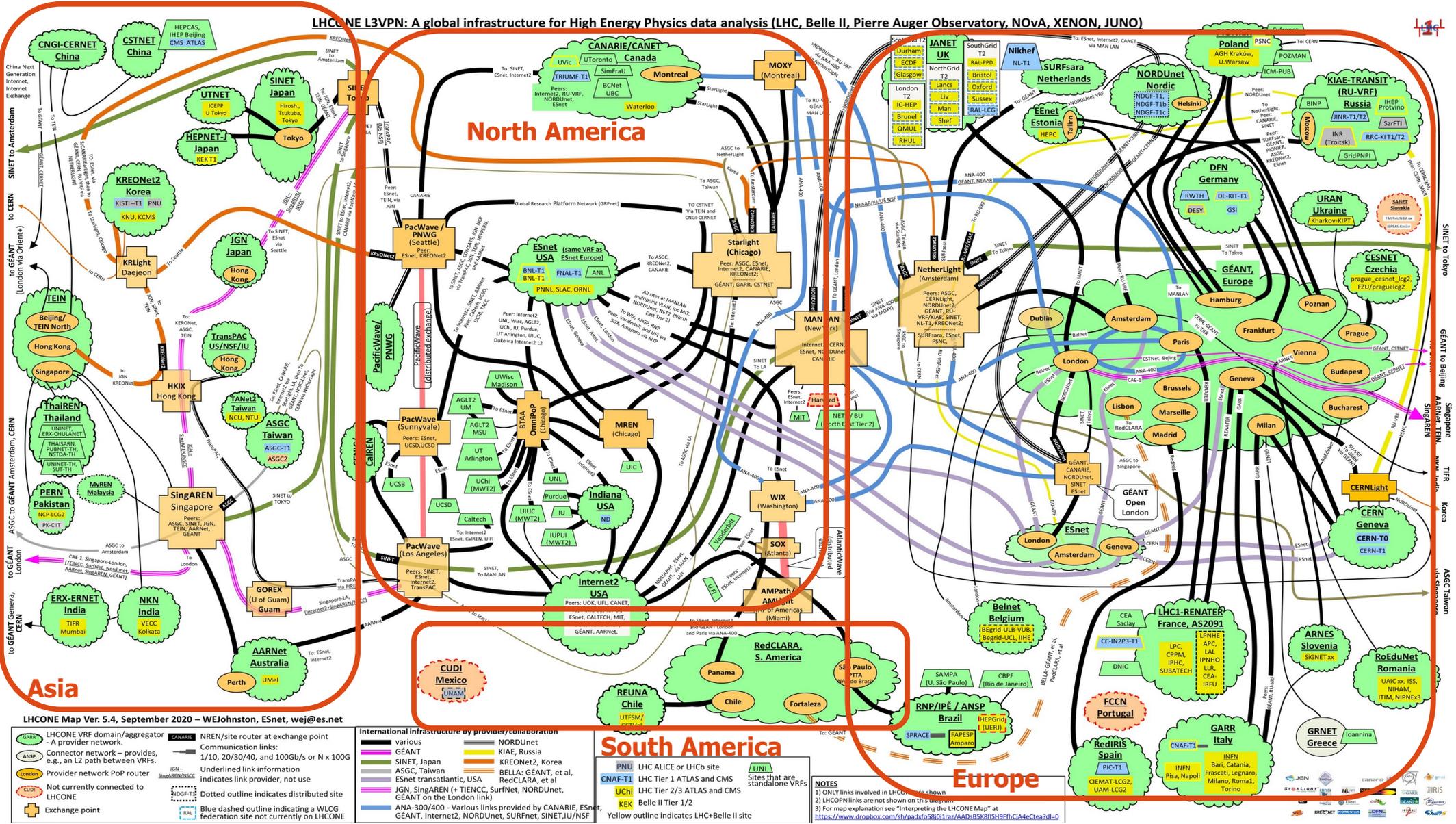


L3VPN status

- **VRFs: 30 national and international Research Networks**
- **Connected sites: ~110**
- Trans-Atlantic connectivity provided by ESnet, GEANT, Internet2, NORDUnet and SURFnet
- Trans-Pacific connectivity provided by ASGCnet, KREOnet, SINET, TransPAC
- Interconnections at Open Exchange Points including NetherLight, StarLight, MANLAN, WIX, CERNlight and others



LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NoVA, XENON, JUNO)



North America

South America

Europe

Asia

LHCONE Map Ver. 5.4, September 2020 – WEJohnston, ESnet, wej@es.net

- LHCONE VRF domain/aggregator**
 - Green circle: LHCONE VRF domain/aggregator
 - Black circle: A provider network
 - Orange circle: Connector network – provides, e.g., an L2 path between VRFs
 - Blue circle: Provider network PoP router
 - Grey circle: Not currently connected to LHCONE
 - Yellow circle: Exchange point
- International infrastructure by provider/collaboration**
 - Black line: various
 - Green line: GEANT
 - Blue line: SINET, Japan
 - Red line: ESnet transatlantic, USA
 - Purple line: JGN, SingAREN (+ TIENCC, SurfNet, NORDUnet, GEANT on the London link)
 - Orange line: ANA-300/400 - Various links provided by CANARIE, ESnet, GEANT, Internet2, NORDUnet, SURFnet, SINET, UJ/NSF
 - Yellow line: NORDUnet
 - Light blue line: KIAE, Russia
 - Dark blue line: KREONet2, Korea
 - Light green line: BELLA; GEANT, et al
 - Light purple line: RedCLARA, et al
 - Light orange line: JGN, SingAREN (+ TIENCC, SurfNet, NORDUnet, GEANT on the London link)
 - Light blue line: ANA-300/400 - Various links provided by CANARIE, ESnet, GEANT, Internet2, NORDUnet, SURFnet, SINET, UJ/NSF
- Other symbols**
 - Black square: NREN/site router at exchange point
 - Black line: Communication links; 1/10, 20/30/40, and 100Gb/s or N x 100G
 - Underlined link information: indicates link provider, not use
 - Dotted outline: indicates distributed site
 - Blue dashed outline: indicating a WLCG federation site not currently on LHCONE

- Notes**
 - 1) ONLY links involved in LHCONE are shown
 - 2) LHCOPN links are not shown on this diagram
 - 3) For map explanation see "Interpreting the LHCONE Map" at <https://www.dropbox.com/sh/paxifos8017raz/AADt85K8fH9fHf4A4Ctea7d?dl=0>
- Legend for South America**
 - PNU: LHC ALICE or LHCb site
 - LANL: LHC Tier 1 ATLAS and CMS
 - UCHI: LHC Tier 2/3 ATLAS and CMS
 - KEK: Belle II Tier 1/2
 - Sites that are standalone VRFs
 - Yellow outline: indicates LHC+ Belle II site

Yellow outline indicates LHC+ Belle II site

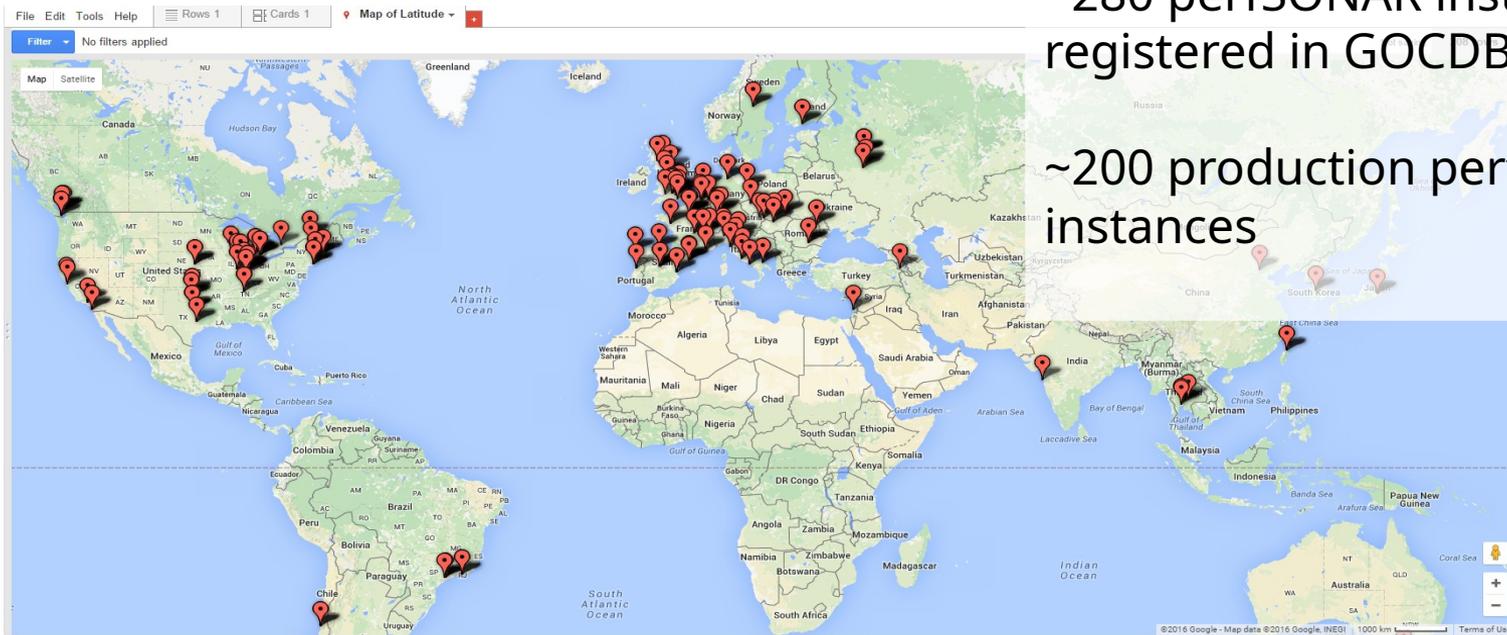
Notes:
 1) ONLY links involved in LHCONE are shown
 2) LHCOPN links are not shown on this diagram
 3) For map explanation see "Interpreting the LHCONE Map" at <https://www.dropbox.com/sh/paxifos8017raz/AADt85K8fH9fHf4A4Ctea7d?dl=0>



Open to other HEP collaborations



Monitoring: perfSONAR



~280 perfSONAR instances registered in GOCDB/OIM

~200 production perfSONAR instances

- Initial deployment coordinated by WLCG perfSONAR TF
- Commissioning of the network followed by WLCG Network and Transfer Metrics WG

Slide credit: Shawn McKee, University of Michigan

Monitoring: perfSONAR

Some useful dashboards:

- WLCG MadDASH
- Latency per area
- Throughput
- end2end performance
- IPv4 vs IPv6

perfSONAR



Monitoring: Looking Glass

Looking-glass to analyse the routing tables of the VRFs

Peering with these VRFs:

- ASGC AS24167
- CERNlight AS20641
- KREOnet AS17579
- GEANT AS20965
- CANARIE AS6509
- ESnet AS293
- NORDUnet AS2603
- RU-VRF AS57484

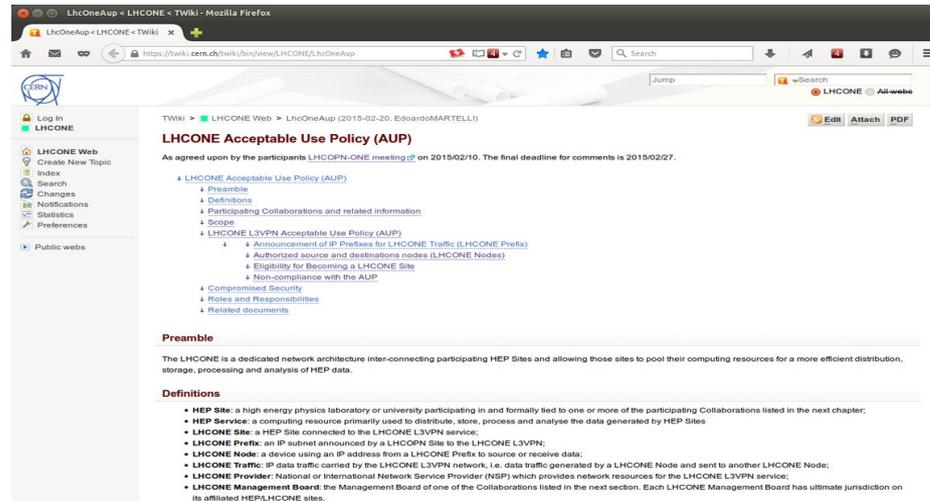
Link: <http://lhcone-lg.cern.ch/>



LHCONE Acceptable Use Policy

The LHCONE AUP has been defined to regulate the utilization of the L3VPN service.

- Updated at the LHCONE meeting of March 2021
- <https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneAup>



The screenshot shows a web browser displaying the LHCONE Acceptable Use Policy (AUP) page. The page title is "LHCONE Acceptable Use Policy (AUP)". The content includes a preamble, definitions, and a list of key terms. The definitions section lists:

- **HEP Site:** a high energy physics laboratory or university participating in and formally tied to one or more of the participating Collaborations listed in the next chapter;
- **HEP Service:** a computing resource primarily used to distribute, store, process and analyse the data generated by HEP Sites;
- **LHCONE Site:** a HEP Site connected to the LHCONE L3VPN service;
- **LHCONE Prefix:** an IP subnet announced by a LHCONE Site to the LHCONE L3VPN;
- **LHCONE Node:** a device using an IP address from a LHCONE Prefix to source or receive data;
- **LHCONE Traffic:** IP data traffic carried by the LHCONE L3VPN network, i.e. data traffic generated by a LHCONE Node and sent to another LHCONE Node;
- **LHCONE Provider:** National or International Network Service Provider (NSP) which provides network resources for the LHCONE L3VPN service;
- **LHCONE Management Board:** the Management Board of one of the Collaborations listed in the next section. Each LHCONE Management Board has ultimate jurisdiction on its affiliated HEP/LHCONE sites.



Database for network information



Database:

- CRIC (Computing Resources Information Catalogue) is the database used by the WLCG sites to document computing resources
- Network information about LHCOPN, LHCONE and related monitoring have been added to CRIC

Tools:

- A WHOIS route-set RS-LHCONE is automatically updated from CRIC
- The route-set can be used by NRENs and sites to build security routing filters

Status:

- schema and web interfaces implemented
- database populated from LHCONE routing tables and reviewed by site managers
- monitoring information being added to CRIC



More info here https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneVRF#Connected_Sites

WLCG data challenges

HL-LHC requirements for Run4 (2029)

The computing model of the experiments at HL-LHC will be different from what they have today: **ATLAS and CMS will produce**

350PB/year/experiment, to be exported in real time to Tier1s

- This would require **4.8Tbps from CERN to the Tier1s**, of which 1.25Tbps over the Atlantic
- **Larger Tier1s are supposed to get connected to CERN and to their Tier2s at 1Tbps** (1Tbps in from Tier0, 1Tbps out to Tier2s)
- Based on these targets, a plan for data challenges is being proposed for the years preceding Run4. First challenge successfully completed in 2021
- Data challenges will use the production infrastructure and will co-exist with production activities.

Data and Network challenges

Network challenges consist in **demonstrating the capability to transfer an increasing volume of data over the years to reach the production transfer target**, sustained for a few days, by the start of HL-LHC in 2029.

Foreseen **milestones at 15% of the target in 2021 (accomplished), 35% in 2023-24, 60% in 2026-27 and 100% in 2029.**

Dates could be adjusted based on the growth plan of the NRENs

R&D activities

Research Network Technology WG

Working on network research projects, with contributions not only from WLCG and NREN, but also from outside, like RFC editors, Linux kernel developers

Publishing results at new web site scitags.org

Packet marking activity: focusing on IPv6 flowlabel field and UDP Firefly packets

- Proposed a packet marking schema for IPv6 flowlabel field
- Flowlabel marking being tested and will be demonstrated at SC22
- Accounting possible with P4 programmable switches

BitPattern	ScienceDomain	Application	Hdr Bit 12	Hdr Bit 13	Hdr Bit 14	Hdr Bit 15	Hdr Bit 16	Hdr Bit 17	Hdr Bit 18	Hdr Bit 23	Hdr Bit 24	Hdr Bit 29	Hdr Bit 30	Hdr Bit 31
xx10000000x000001xx	ATLAS	perfSONAR	x	x	1	0	0	0	0	x	0	1	x	x
xx01000000x000001xx	CMS	perfSONAR	x	x	0	1	0	0	0	x	0	1	x	x
xx11000000x000001xx	LHCb	perfSONAR	x	x	1	1	0	0	0	x	0	1	x	x
xx00100000x000001xx	ALICE	perfSONAR	x	x	0	0	1	0	0	x	0	1	x	x
xx10100000x000001xx	BelleII	perfSONAR	x	x	1	0	1	0	0	x	0	1	x	x
xx01100000x000001xx	SKA	perfSONAR	x	x	0	1	1	0	0	x	0	1	x	x
xx11100000x000001xx	LSST	perfSONAR	x	x	1	1	1	0	0	x	0	1	x	x
xx00010000x000001xx	DUNE	perfSONAR	x	x	0	0	0	1	0	x	0	1	x	x

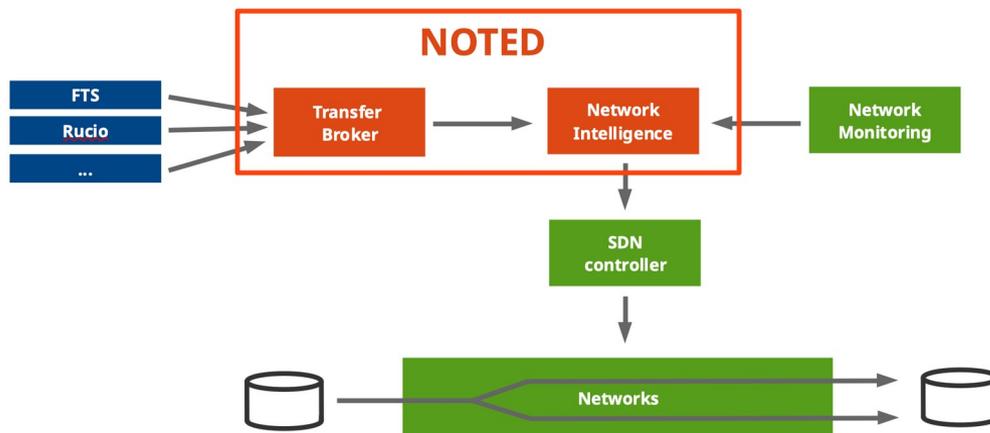
NOTED

NOTED is a framework that can detect large FTS data transfers and trigger network optimization actions to speed up the transfers executions

Already tested with production transfers:

- CERN-PIC with LHCOPN-LHCONE load balancing
- CERN-TRIUMF and KIT-TRIUMF with the activation of dynamic circuits

A new demo planned for SC22

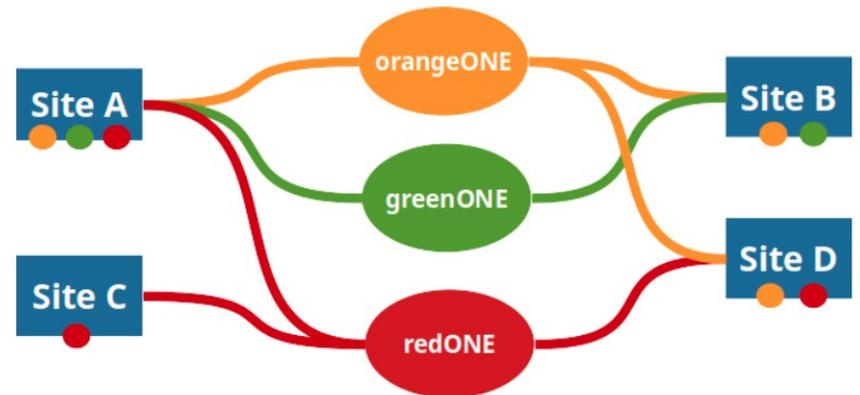


MultiONE

LHCONE's main advantage is threatened by its own success: the more sites connect, the less secure it becomes

A solution would be to create multiple VPN, one for each sub-groups (e.g. ATLAS, CMS, BelleII..). But it is difficult to separate the traffic for sites that participate to multiple VPNs

A solution based on the flowlabel tagging is being explored



Conclusions

Summary

- LHCOPN: upgrading links to CERN to 100Gbps and 400Gbps for Run3
- LHCONE:
 - NRENs upgrading global infrastructure
 - adopted CRIC for network resource description
 - improving monitoring capabilities
- WLCG is running data challenges in preparation of Run4. The LHCONE community will contribute and support them
- On-going research activities to fix weaknesses

References

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

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

Questions?

edoardo.martelli@cern.ch

