# IPv6 in the WLCG

Duncan Rand
Imperial College London & Jisc

Tim Chown
Jisc

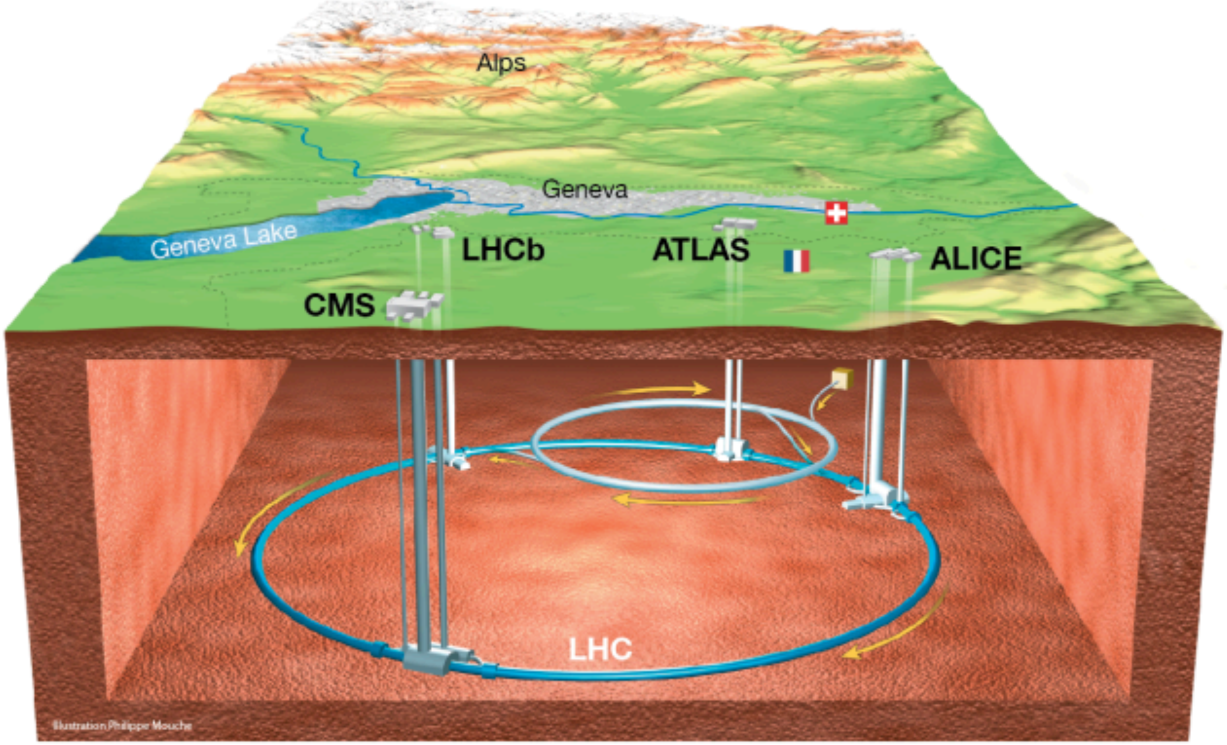
Belgian IPv6 Council Meeting, Brussels, June 2019

# Contents

- The Large Hadron Collider (LHC)
- The Worldwide LHC Computing Grid (WLCG)
- GridPP
- Why IPv6?
- IPv6 deployment plan
- Status of the WLCG sites
- Data transfer with the File Transfer Service (FTS)
- Network monitoring with perfSONAR
- Summary

# The Large Hadron Collider (LHC)

- The LHC is located at CERN on the Franco-Swiss border
- Proton proton and heavy ion collider with four main experiments
- Two general purpose: ATLAS and CMS
- Two specialist: LHCb and ALICE (heavy ions)
- During Run 1 at 8 TeV: found the Higgs particle in 2012
- Started Run 2 in 2015 at 13 TeV, just finished it on Monday 3rd December
- Computing for LHC experiments carried out by the Worldwide LHC Computing Grid (WLCG or ‘the Grid’)

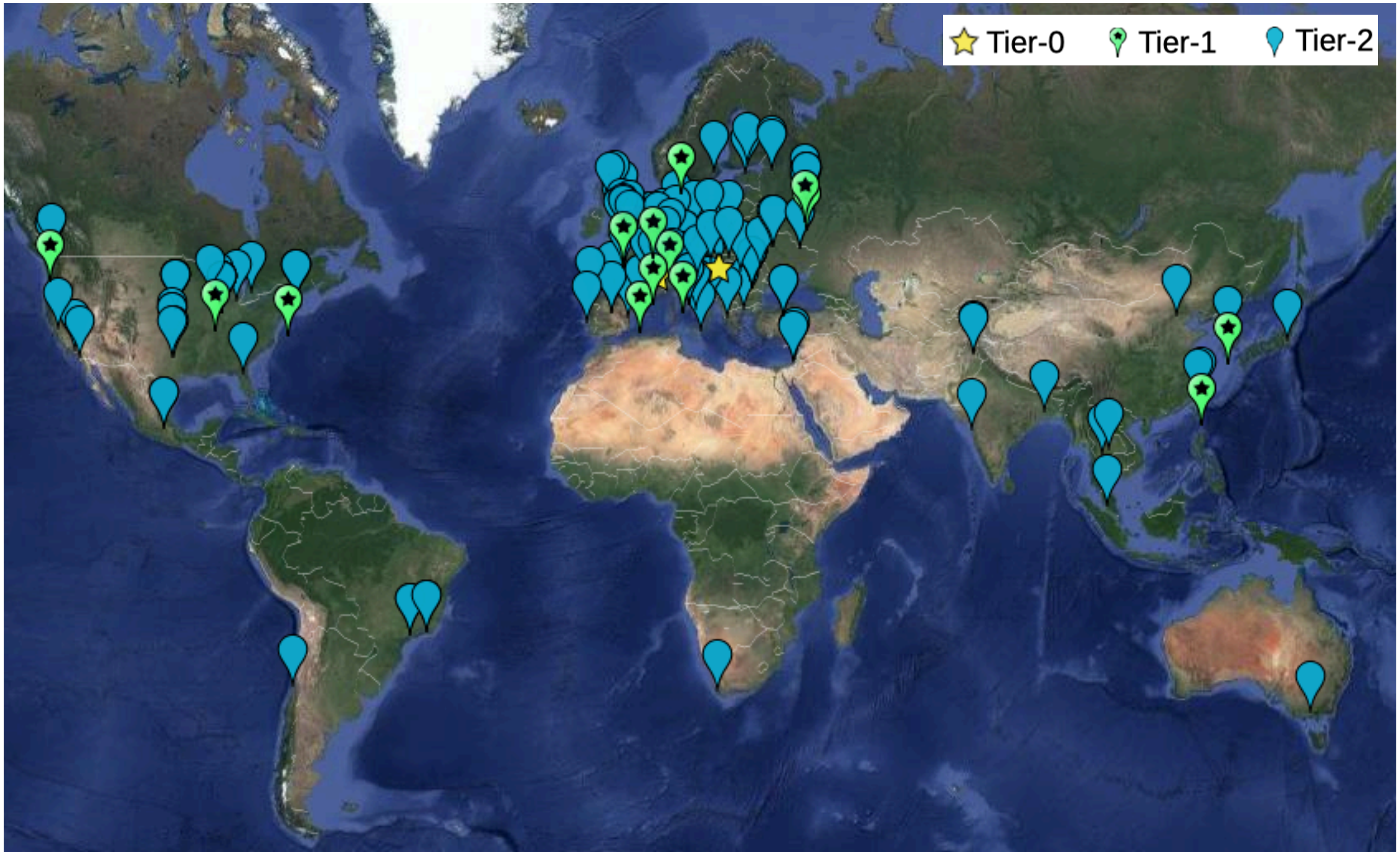


**© 2014-2018 CERN**

# Worldwide LHC Computing Grid (WLCG)

- The WLCG is a global collaboration of more than 170 computing centres in 42 countries.
- Its mission is to provide global computing resources to store, distribute and analyse the ~50-70 petabytes of data generated per year by the LHC experiments

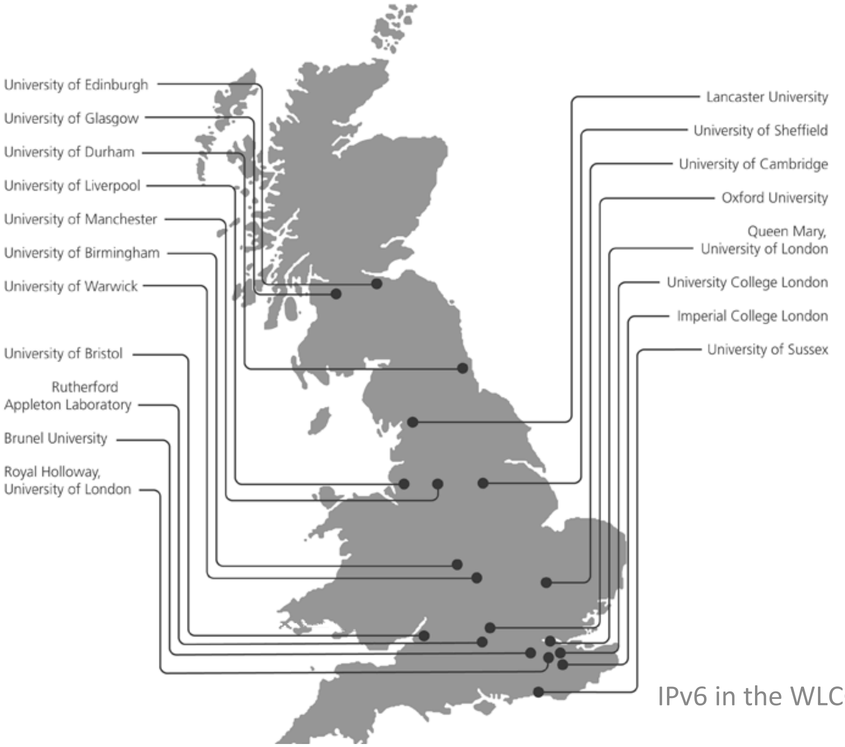
- Sites hierarchically arranged
- Tier-0 at CERN (and Wigner in Hungary)
- 14 Tier-1s (mainly national laboratories)
- 149 Tier-2s (generally university physics laboratories)

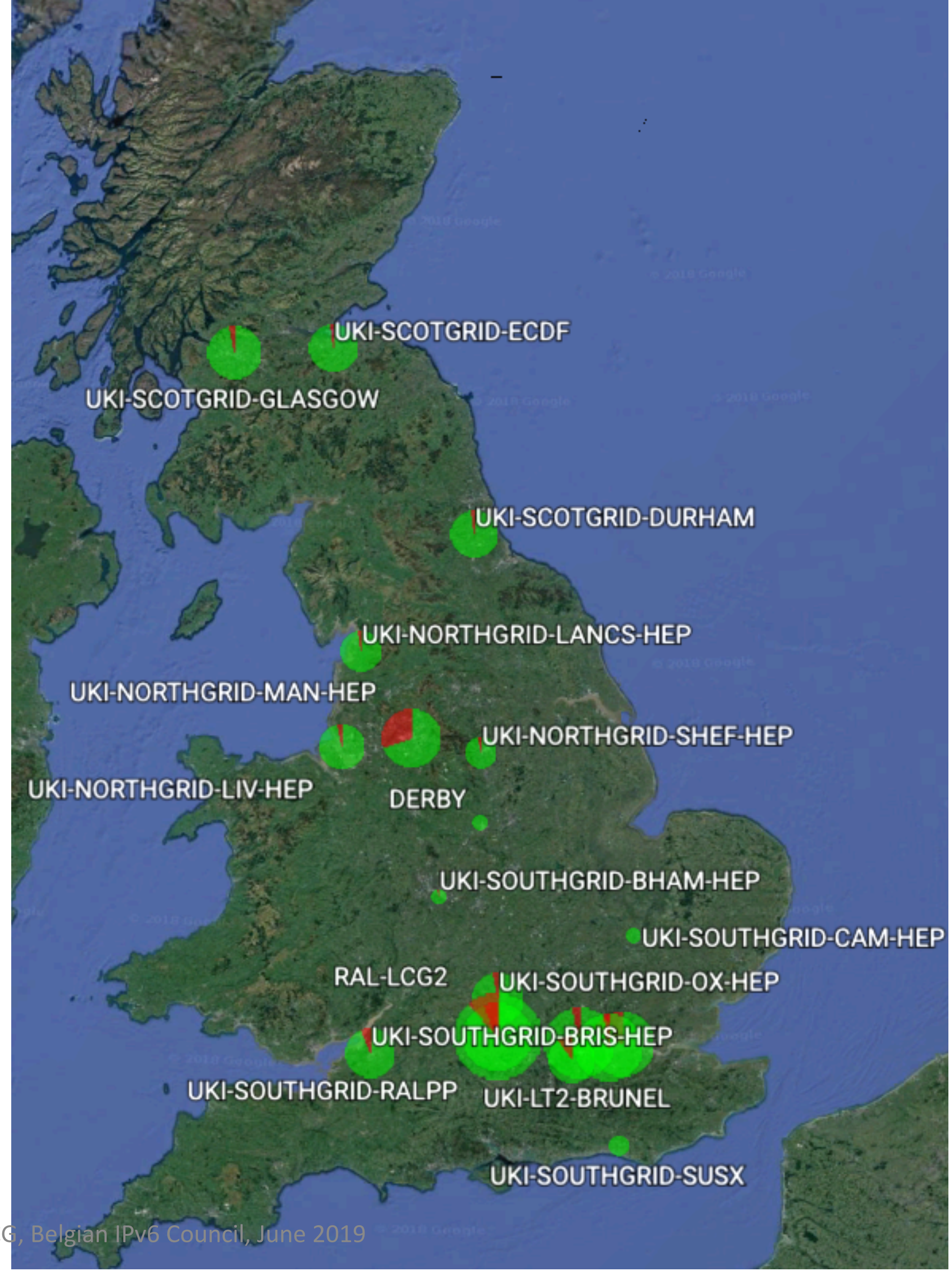


GridPP

UK Computing for Particle Physics

GridPP is a collaboration of nineteen institutes providing data-intensive distributed computing resources for the UK High Energy Physics community and the UK contribution to the WLCG





# WLCG Tiers Hierarchy

- Initial modelling of LHC computing requirements suggested a hierarchical tier-based data management and transfer model
- Data exported from Tier-0 at CERN to each Tier-1 and then on to Tier-2s
- However better than expected network bandwidth means that the LHC experiments have been able to relax this hierarchy
- Now data is transferred in an all-to-all mesh configuration
- Data often transferred across multiple domains
- e.g. a CMS transfer to Imperial College London might come from Fermilab near Chicago

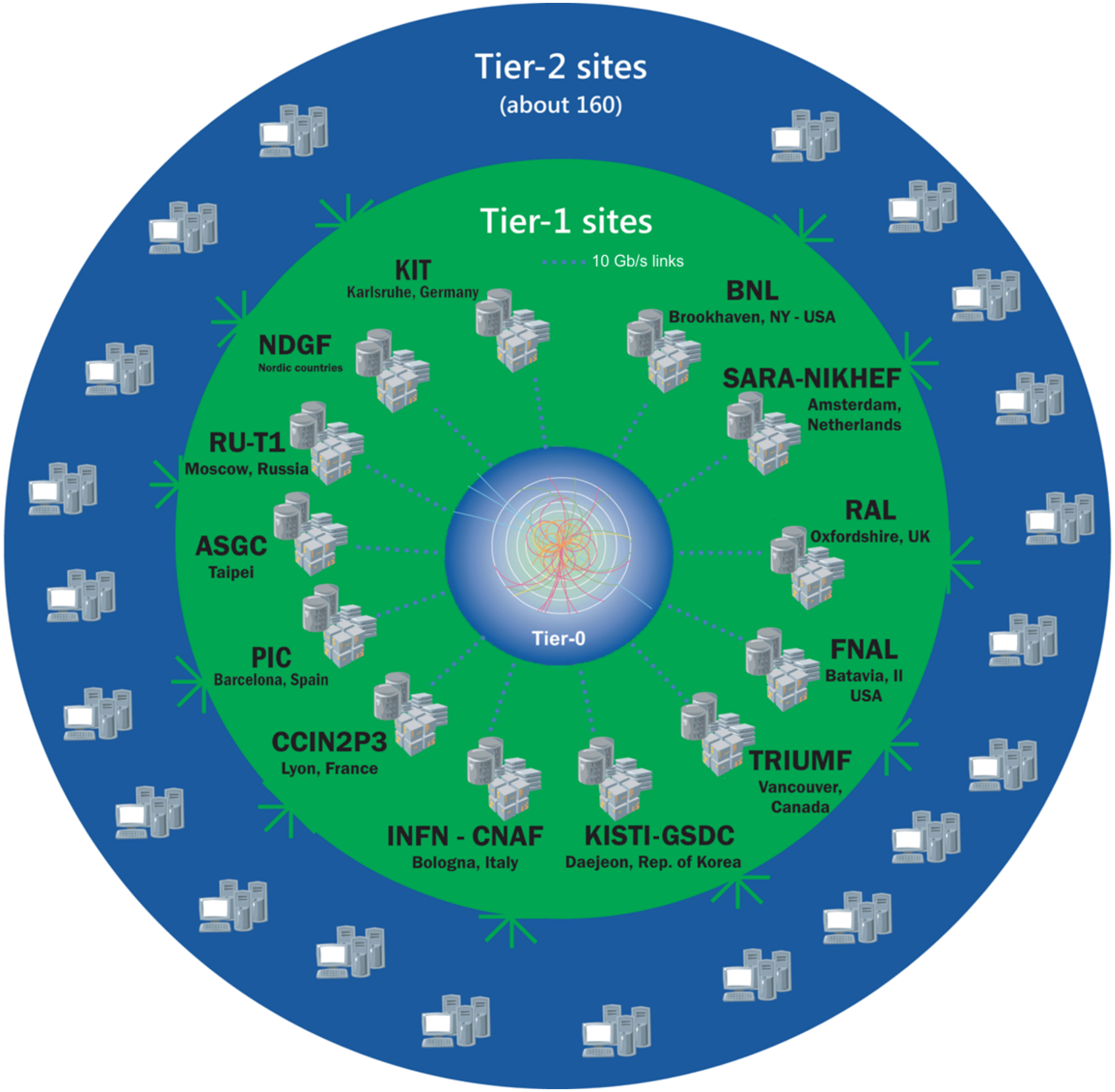


Image from 2014

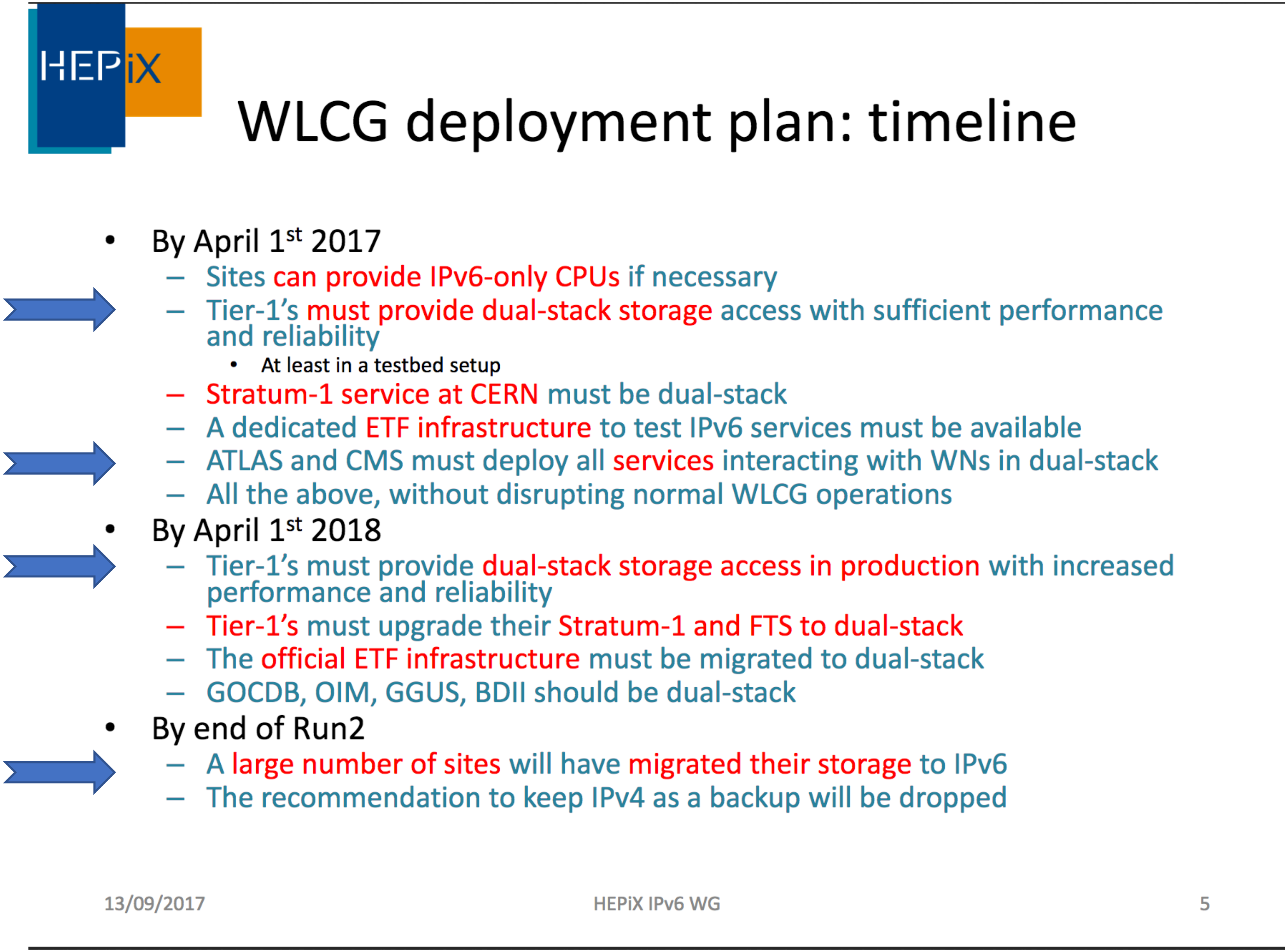
# WLCG Site Operation

- WLCG resources at a site generally consist of
  - a large compute cluster (typically several thousand cores)
  - a disk storage cluster (typically a few petabytes)
- Bulk data is transferred between storage clusters with the File Transfer Service (FTS3) using GridFTP
- Computing jobs arrive at the site and produce simulated data or process some of the data stored locally
- Also possible for a job at one site to access data directly from storage at another
- For example QMUL reads CMS data from storage at Imperial College
- It is envisaged that the use of such remote reading of data is likely to increase in the future

# Why IPv6?

- The WLCG is generally open to new compute resources
- We might get an offer of opportunistic CPU resources which are IPv6-only – want to be able to use them
- So the main goal is to make the data at the sites accessible by clients running on IPv6-only machines
- Also for pledged resources, sites running out of IPv4 addresses and to avoid use of NAT

- Initial deployment plan
  - Make experiment central services dual-stack
  - Make some test worker nodes IPv6-only
  - Deploy perfSONAR network monitoring at sites
  - Make site storage accessible over IPv6

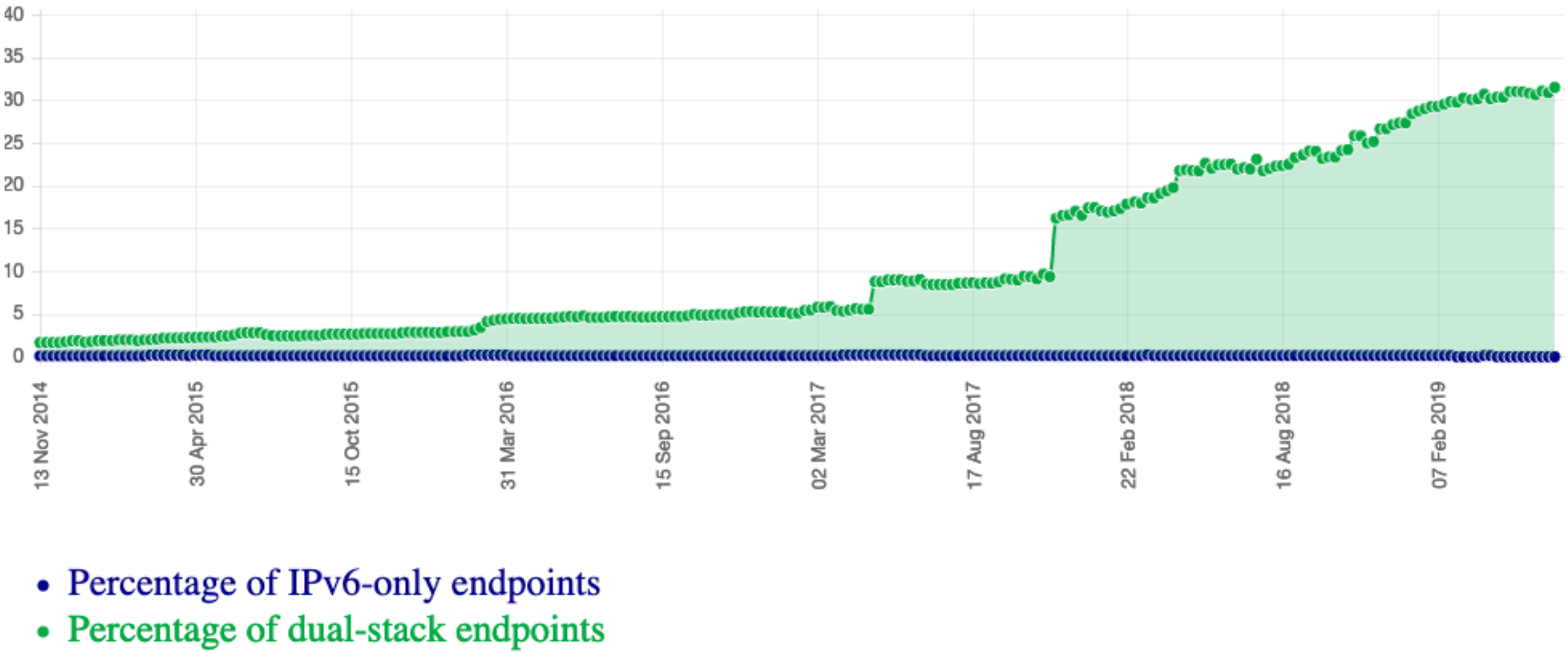


# WLCG deployment plan: timeline

- By April 1st 2017
  - Sites can provide IPv6-only CPUs if necessary
  - Tier-1's must provide dual-stack storage access with sufficient performance and reliability
    - At least in a testbed setup
  - Stratum-1 service at CERN must be dual-stack
  - A dedicated ETF infrastructure to test IPv6 services must be available
  - ATLAS and CMS must deploy all services interacting with WNs in dual-stack
  - All the above, without disrupting normal WLCG operations
- By April 1st 2018
  - Tier-1's must provide dual-stack storage access in production with increased performance and reliability
  - Tier-1's must upgrade their Stratum-1 and FTS to dual-stack
  - The official ETF infrastructure must be migrated to dual-stack
  - GOCDB, OIM, GGUS, BDII should be dual-stack
- By end of Run2
  - A large number of sites will have migrated their storage to IPv6
  - The recommendation to keep IPv4 as a backup will be dropped

13/09/2017 HEPiX IPv6 WG 5

# Growth of dual-stack  hosts in the WLCG

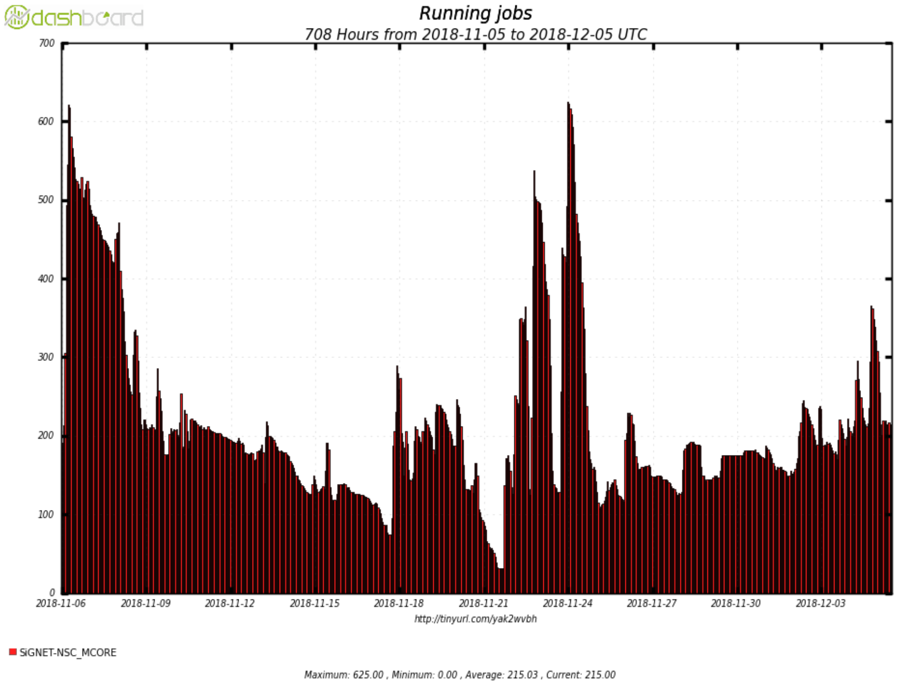


~30%

Fraction of endpoints listed in the CERN central BDII (lcg-bdii.cern.ch) where the DNS returns a dual-stack IPv6-IPv4 (A+AAAA) resolution (green line) or an IPv6-only resolution (blue line). (https://orsone.mi.infn.it/~prelz/ipv6_bdii/).

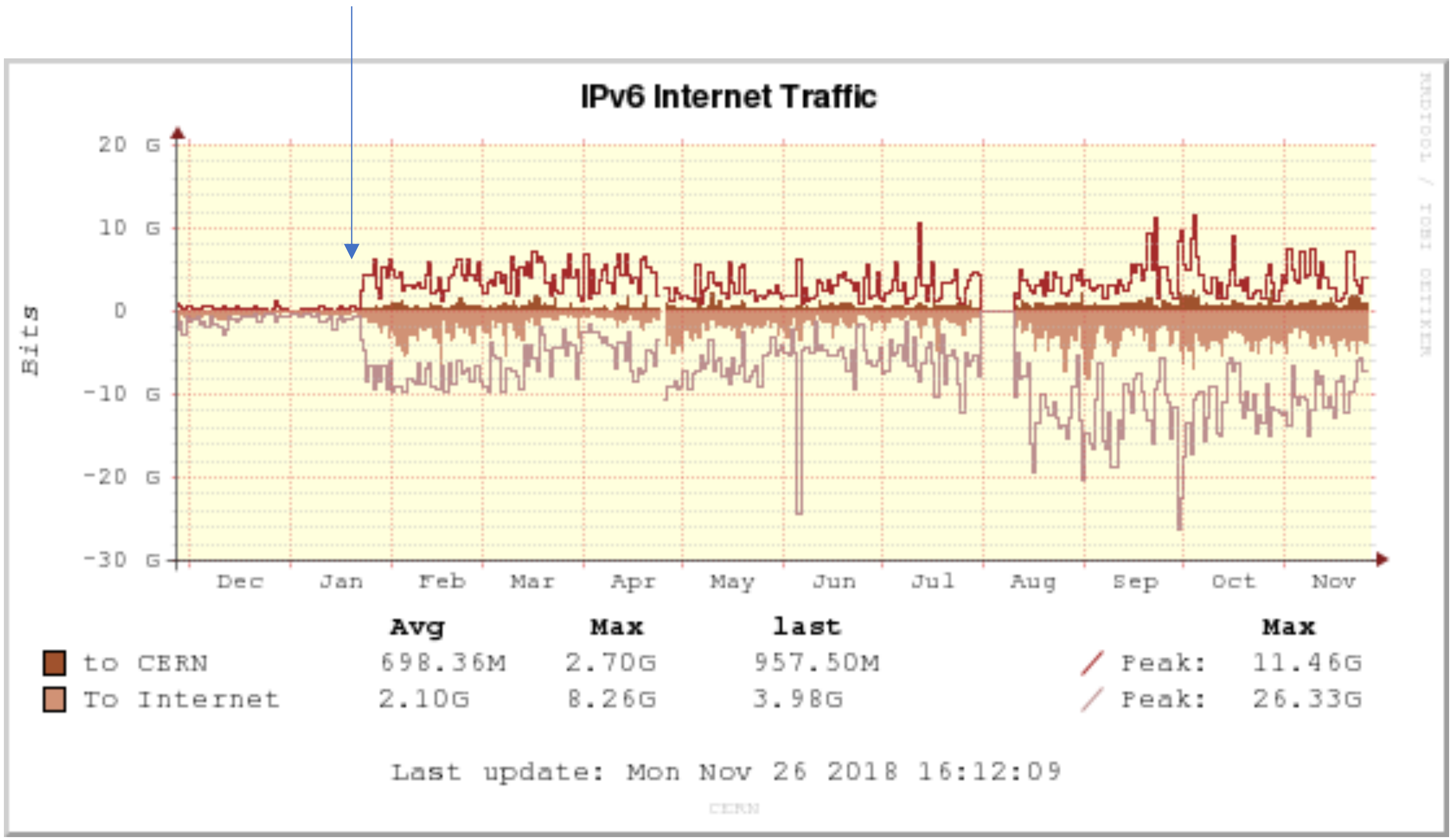
# IPv6-only compute

- Need to be ready for a possible offer of IPv6-only compute resources
- Testing IPv6-only worker nodes for
  - CMS at Brunel University London
  - ATLAS at the Jozef Stefan Institute, Slovenia (running production jobs)



# Turning on IPv6 on CERN Tier-0 disk storage (EOS) in Jan 2018

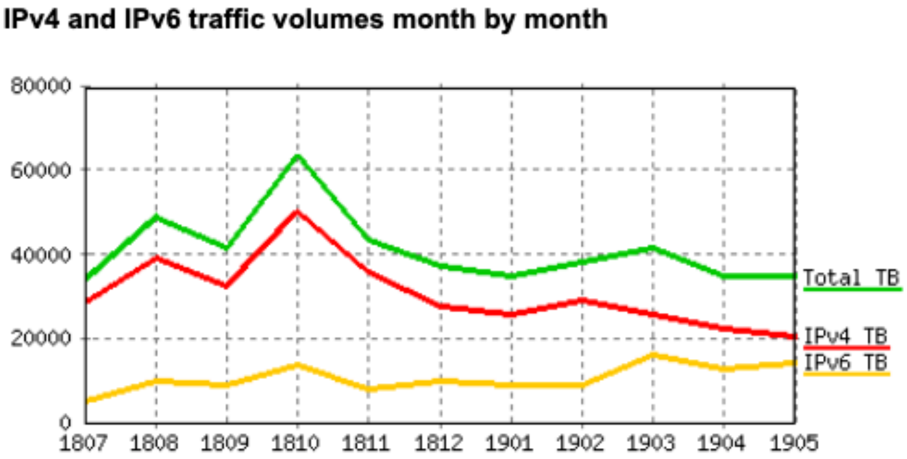
Non-LHCOPN/non-LHCONE traffic

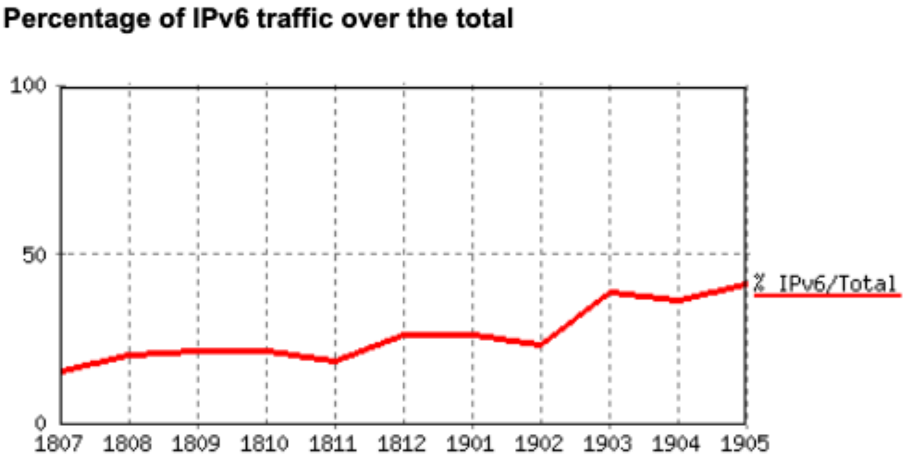


| | Avg | Max | last | | Max |
|---|---|---|---|---|---|
| 🟫 to CERN | 698.36M | 2.70G | 957.50M | / Peak: | 11.46G |
| 🟫 To Internet | 2.10G | 8.26G | 3.98G | / Peak: | 26.33G |

Last update: Mon Nov 26 2018 16:12:09

# LHCOPN and LHCONE IPv4 and IPv6 traffic volumes seen at CERN Tier-0





~41%

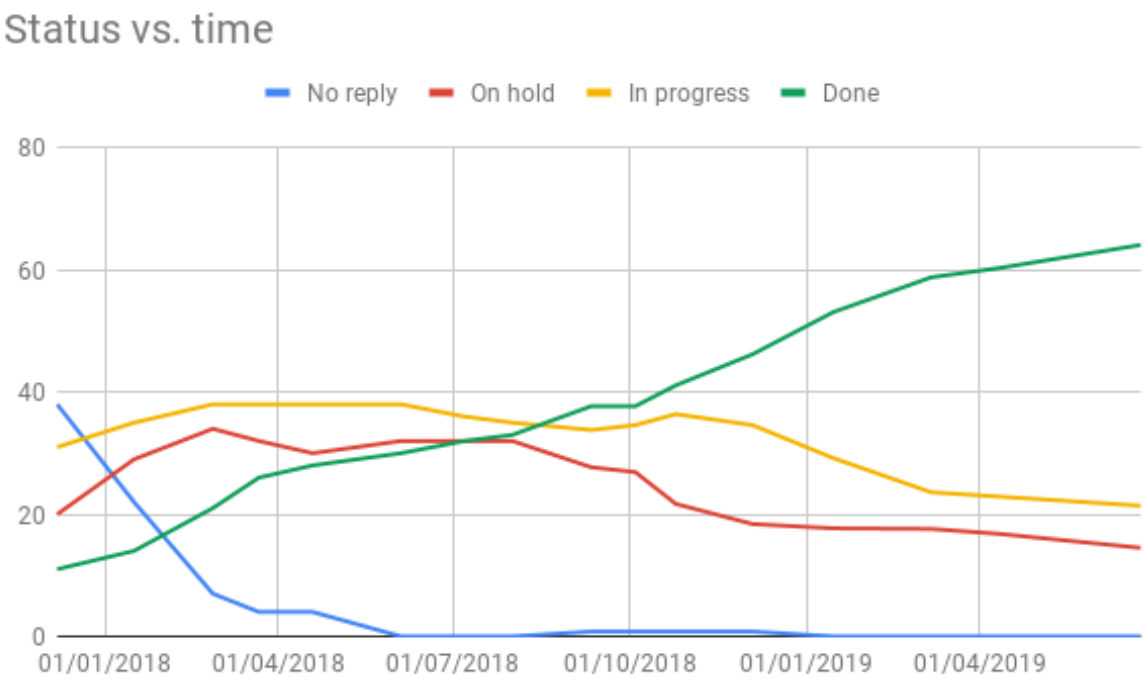
# Tier-1 and Tier-2 dual-stack roll-out

- Fourteen Tier-1s have dual-stack storage and one has IPv4

- Tier-2 sites were requested to deploy dual-stack perfSONAR and storage by end of Run 2 (end of 2018)
- Submitted a ticket to each site in autumn 2017 requesting timescale for deployment of IPv6 and details of steps
- Following up with assistance, checking deployment etc
- Several sites are waiting for their campus network infrastructure to become IPv6-ready
- Only a few sites where the problem is at the Grid service level

# Tier-2 evolution

Tier-2 IPv6 deployment ticket states (%)

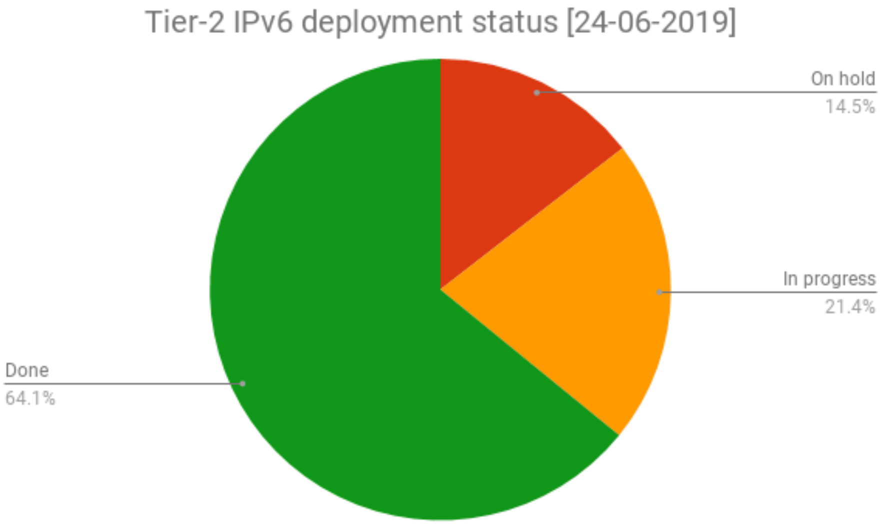


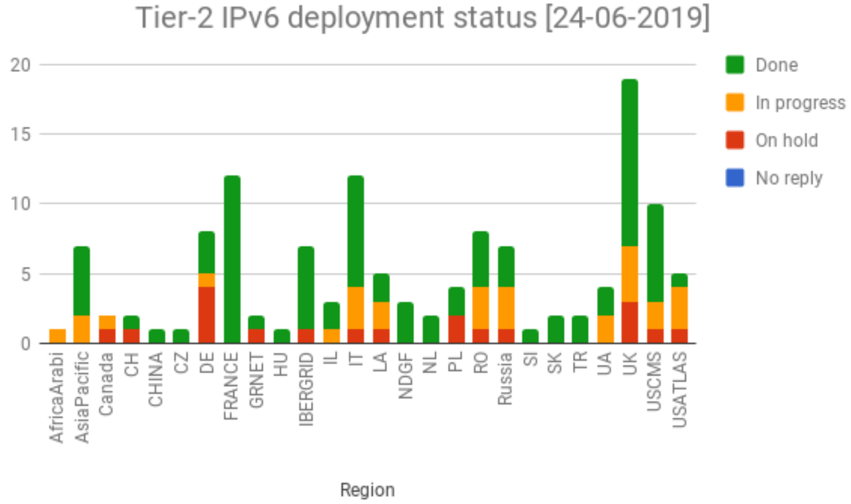
June 2019

# Tier-2s: GGUS tickets submitted to 115 Tier-2 sites

64% Tier-2s with dual-stack perfSONAR and storage

https://twiki.cern.ch/twiki/bin/view/LCG/WlcgIpv6#WLCG_Tier_2_IPv6_deployment_stat



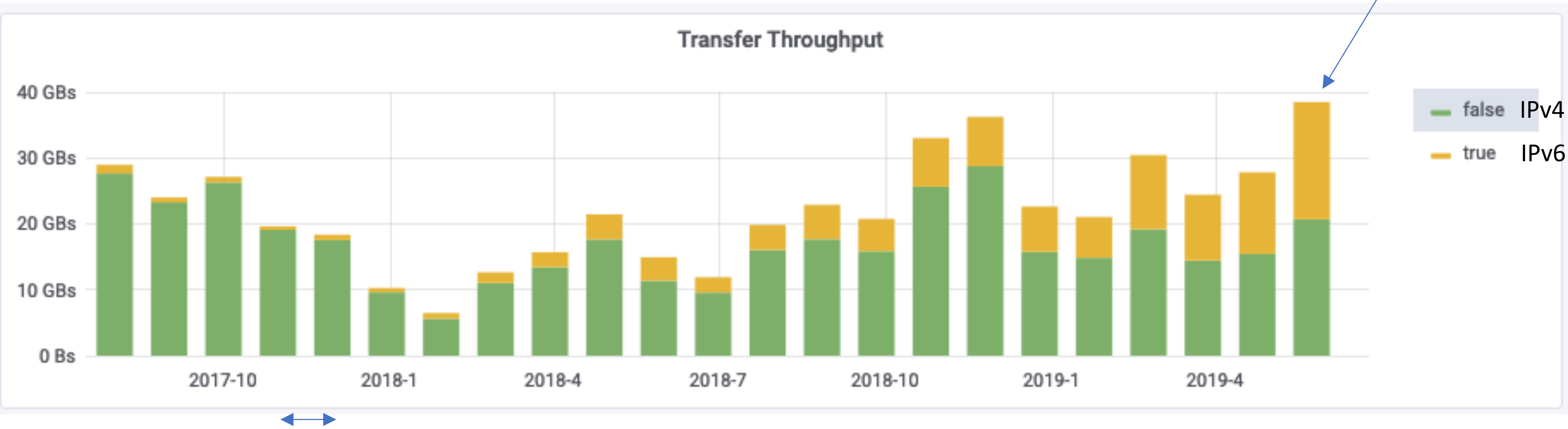


# Proportion of Storage accessible over IPv6  (June 2019)

| Experiment | Fraction of Tier-2 storage accessible via IPv6 |
|---|---|
| ALICE | 81% |
| ATLAS | 57% |
| CMS | 84% |
| LHCb | 69% |
| WLCG overall | 70% |

| Country | Fraction of Tier-2 storage accessible via IPv6 |
|---|---|
| UK (GridPP) | 77% |

# Bulk data transfer using the File Transfer Service (FTS)



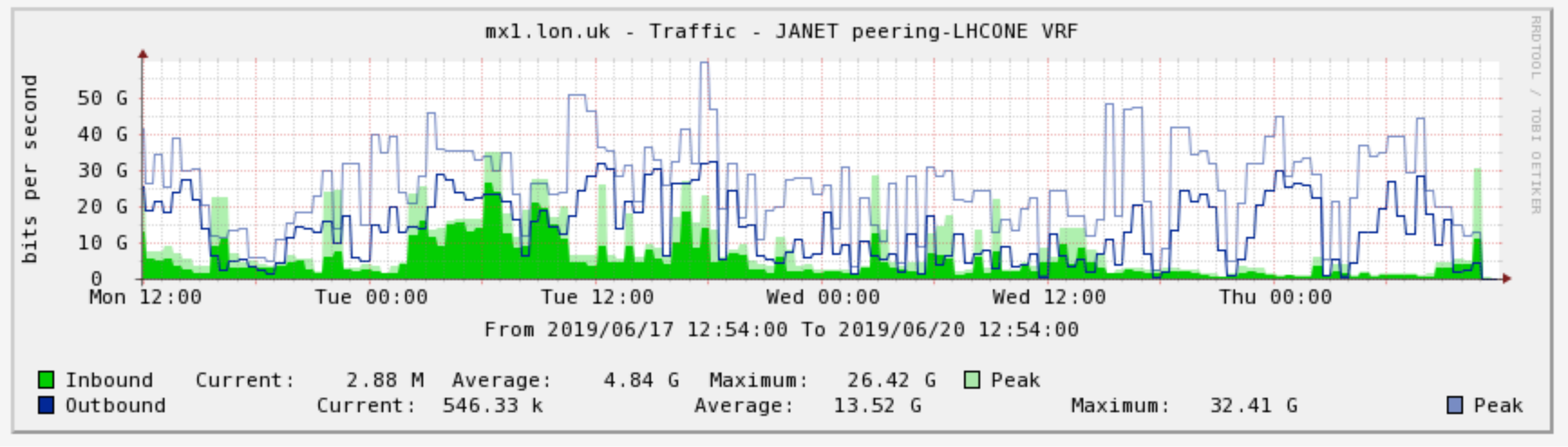
~52% IPv6

IPv4

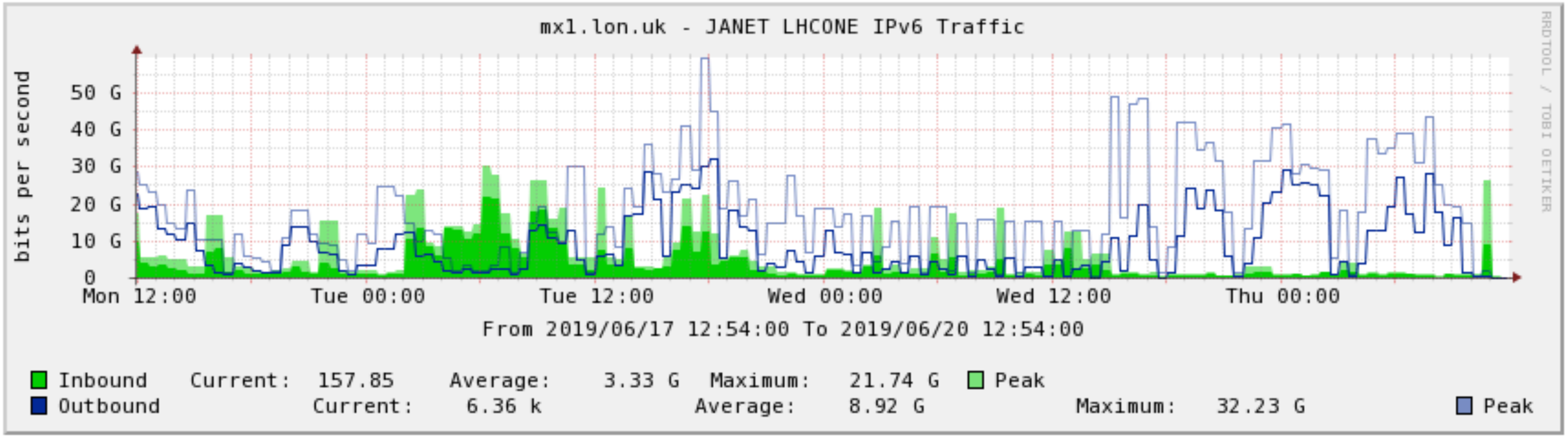
IPv6

Aggregate WLCG transfer rates (gigabytes/s)

# Janet (UK) LHCONE traffic

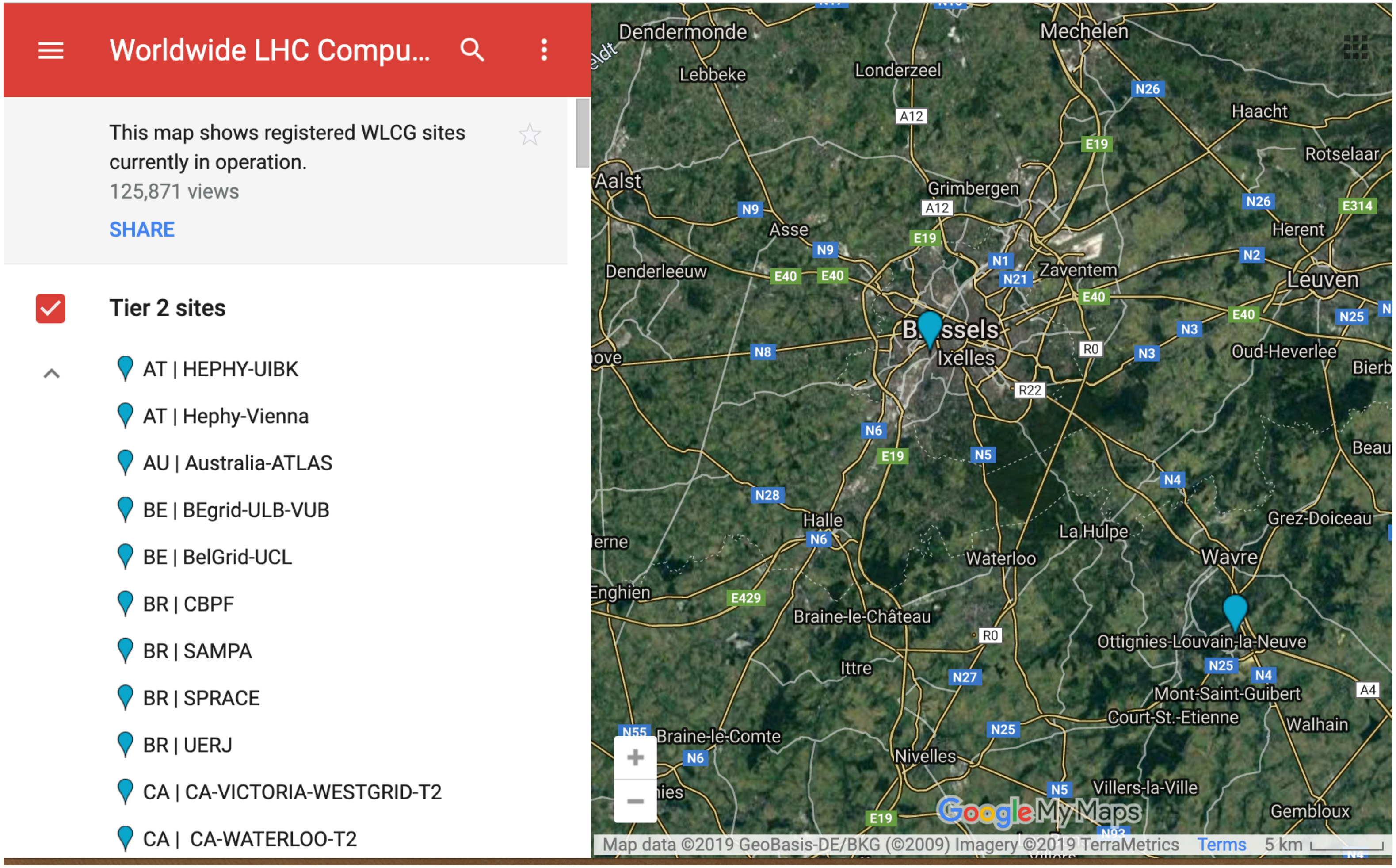


All traffic



IPv6 traffic

https://tools.geant.net/portal/links/p-cacti/graph_view.php?action=tree&tree_id=30&leaf_id=23059



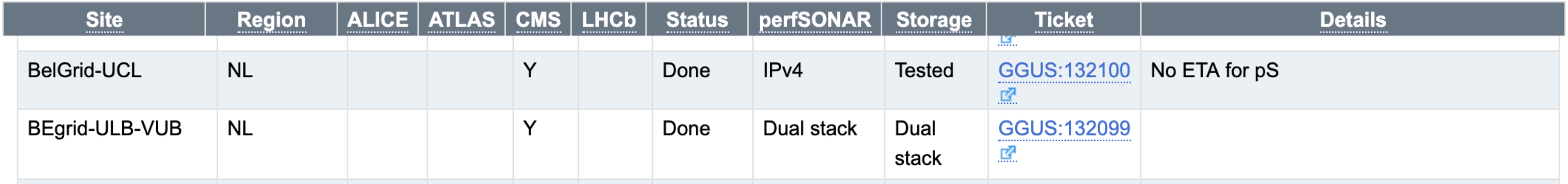
Worldwide LHC Compu...

This map shows registered WLCG sites currently in operation.
125,871 views

SHARE

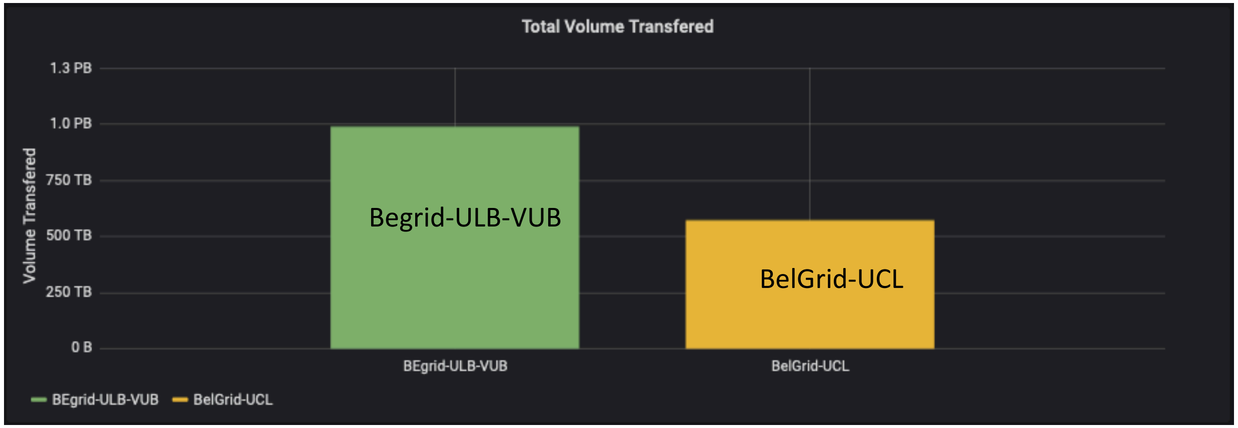
☑ Tier 2 sites

- AT | HEPHY-UIBK
- AT | Hephy-Vienna
- AU | Australia-ATLAS
- BE | BEgrid-ULB-VUB
- BE | BelGrid-UCL
- BR | CBPF
- BR | SAMPA
- BR | SPRACE
- BR | UERJ
- CA | CA-VICTORIA-WESTGRID-T2
- CA | CA-WATERLOO-T2

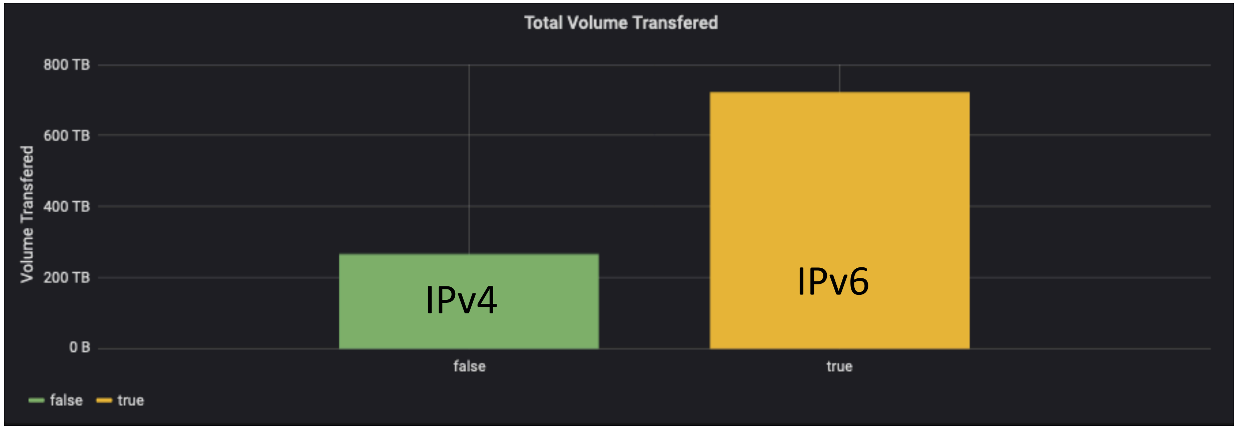


| Site | Region | ALICE | ATLAS | CMS | LHCb | Status | perfSONAR | Storage | Ticket | Details |
|---|---|---|---|---|---|---|---|---|---|---|
| BelGrid-UCL | NL | | | Y | | Done | IPv4 | Tested | GGUS:132100 | No ETA for pS |
| BEgrid-ULB-VUB | NL | | | Y | | Done | Dual stack | Dual stack | GGUS:132099 | |

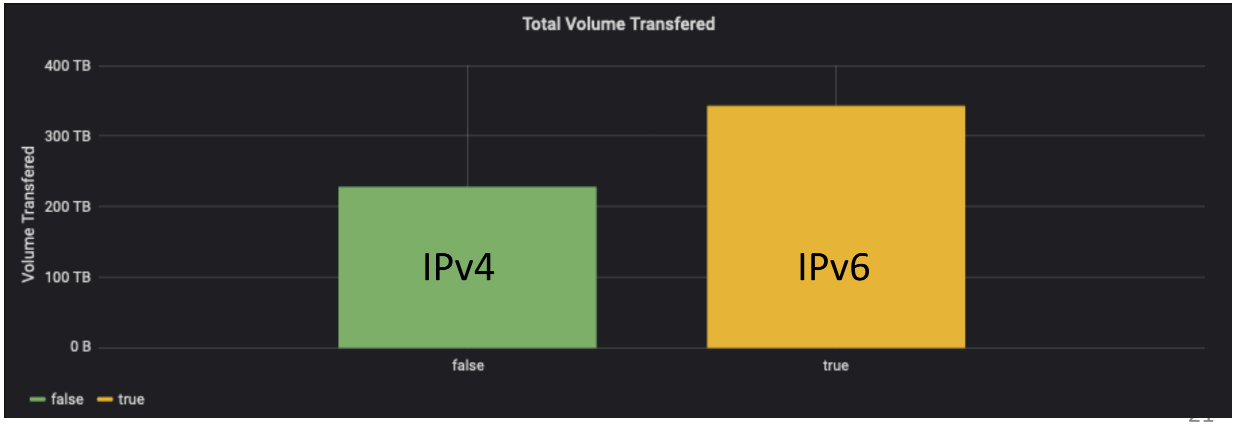
Total volume transferred (FTS, last 30 days)



Begrid-ULB-VUB



BelGrid-UCL



# perfSONAR

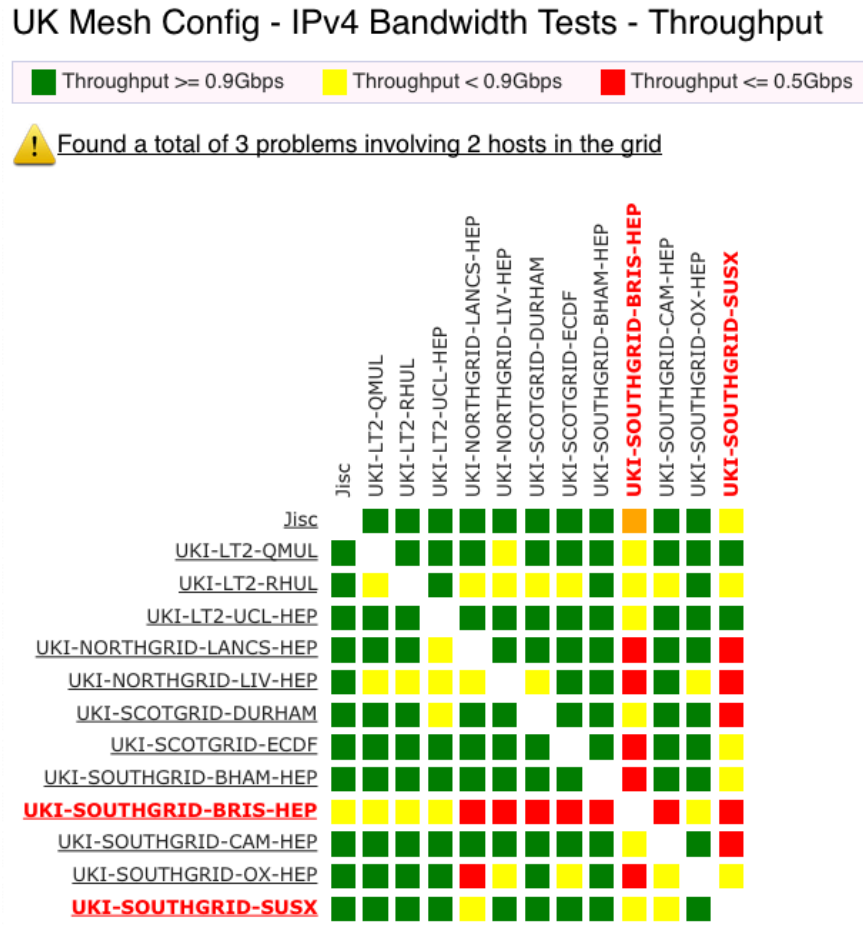
- Network monitoring tool developed by Esnet, GEANT, Indiana University and Internet2
- 'perfSONAR is a widely-deployed test and measurement infrastructure that is used by science networks and facilities around the world to monitor and ensure network performance.'
- http://www.perfsonar.net/about/what-is-perfsonar/

- WLCG goals with perfSONAR
  - Find and isolate “network” problems; alerting in time
  - Characterize network use such as finding base-line performance
  - In the future: provide a source of network metrics for higher level services

- perfSONAR is IPv6 compatible

# perfSONAR dashboards

- Each WLCG site requested to deploy perfSONAR
- WLCG has meshes for a variety of groupings e.g. the LHCOPN, CMS and ATLAS
- UK also runs dual-stack one: throughput, latency, loss, traceroute
- Gives insight into network performance over IPv4 and IPv6 within UK

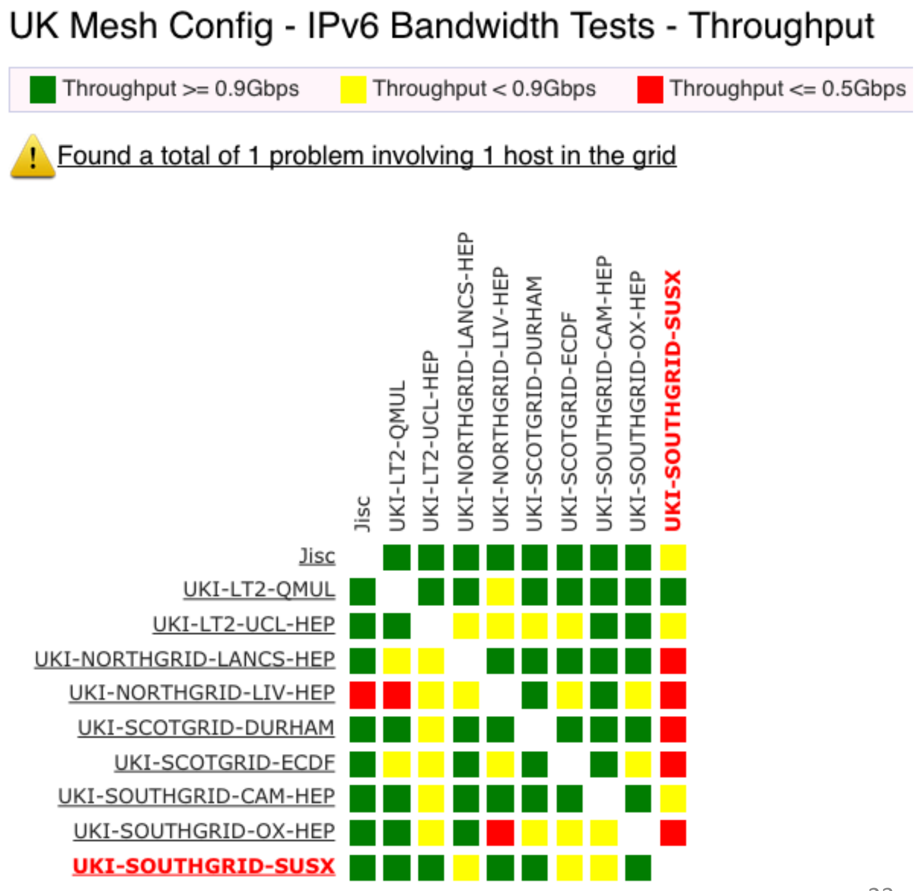


UK Mesh Config - IPv4 Bandwidth Tests - Throughput

🟩 Throughput >= 0.9Gbps 🟨 Throughput < 0.9Gbps 🟥 Throughput <= 0.5Gbps

⚠ Found a total of 3 problems involving 2 hosts in the grid

| | Jisc | UKI-LT2-QMUL | UKI-LT2-RHUL | UKI-LT2-UCL-HEP | UKI-NORTHGRID-LANCS-HEP | UKI-NORTHGRID-LIV-HEP | UKI-SCOTGRID-DURHAM | UKI-SCOTGRID-ECDF | UKI-SOUTHGRID-BHAM-HEP | **UKI-SOUTHGRID-BRIS-HEP** | UKI-SOUTHGRID-CAM-HEP | UKI-SOUTHGRID-OX-HEP | **UKI-SOUTHGRID-SUSX** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jisc | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟧 | 🟩 | 🟩 | 🟨 |
| UKI-LT2-QMUL | 🟩 | | 🟩 | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟩 |
| UKI-LT2-RHUL | 🟩 | 🟨 | | 🟩 | 🟨 | 🟨 | 🟨 | 🟨 | 🟩 | 🟨 | 🟨 | 🟩 | 🟨 |
| UKI-LT2-UCL-HEP | 🟩 | 🟩 | 🟩 | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟩 |
| UKI-NORTHGRID-LANCS-HEP | 🟩 | 🟩 | 🟩 | 🟨 | | 🟩 | 🟩 | 🟩 | 🟩 | 🟥 | 🟩 | 🟩 | 🟥 |
| UKI-NORTHGRID-LIV-HEP | 🟩 | 🟨 | 🟨 | 🟨 | 🟨 | | 🟨 | 🟩 | 🟩 | 🟥 | 🟩 | 🟨 | 🟥 |
| UKI-SCOTGRID-DURHAM | 🟩 | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟥 |
| UKI-SCOTGRID-ECDF | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | | 🟩 | 🟥 | 🟩 | 🟩 | 🟨 |
| UKI-SOUTHGRID-BHAM-HEP | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | | 🟥 | 🟩 | 🟩 | 🟨 |
| **UKI-SOUTHGRID-BRIS-HEP** | 🟨 | 🟨 | 🟨 | 🟨 | 🟥 | 🟥 | 🟥 | 🟥 | 🟥 | | 🟥 | 🟨 | 🟥 |
| UKI-SOUTHGRID-CAM-HEP | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟨 | | 🟩 | 🟥 |
| UKI-SOUTHGRID-OX-HEP | 🟩 | 🟩 | 🟩 | 🟩 | 🟥 | 🟨 | 🟩 | 🟨 | 🟩 | 🟥 | 🟨 | | 🟨 |
| **UKI-SOUTHGRID-SUSX** | 🟩 | 🟩 | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟩 | 🟩 | 🟨 | 🟨 | 🟩 | |



UK Mesh Config - IPv6 Bandwidth Tests - Throughput

🟩 Throughput >= 0.9Gbps 🟨 Throughput < 0.9Gbps 🟥 Throughput <= 0.5Gbps

⚠ Found a total of 1 problem involving 1 host in the grid

| | Jisc | UKI-LT2-QMUL | UKI-LT2-UCL-HEP | UKI-NORTHGRID-LANCS-HEP | UKI-NORTHGRID-LIV-HEP | UKI-SCOTGRID-DURHAM | UKI-SCOTGRID-ECDF | UKI-SOUTHGRID-CAM-HEP | UKI-SOUTHGRID-OX-HEP | **UKI-SOUTHGRID-SUSX** |
|---|---|---|---|---|---|---|---|---|---|---|
| Jisc | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟨 |
| UKI-LT2-QMUL | 🟩 | | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 |
| UKI-LT2-UCL-HEP | 🟩 | 🟩 | | 🟨 | 🟨 | 🟨 | 🟨 | 🟩 | 🟩 | 🟨 |
| UKI-NORTHGRID-LANCS-HEP | 🟩 | 🟨 | 🟨 | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟥 |
| UKI-NORTHGRID-LIV-HEP | 🟥 | 🟥 | 🟨 | 🟨 | | 🟩 | 🟨 | 🟩 | 🟨 | 🟥 |
| UKI-SCOTGRID-DURHAM | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | | 🟩 | 🟩 | 🟩 | 🟥 |
| UKI-SCOTGRID-ECDF | 🟩 | 🟨 | 🟨 | 🟩 | 🟨 | 🟩 | | 🟩 | 🟨 | 🟥 |
| UKI-SOUTHGRID-CAM-HEP | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟩 | 🟩 | | 🟩 | 🟨 |
| UKI-SOUTHGRID-OX-HEP | 🟩 | 🟩 | 🟨 | 🟩 | 🟥 | 🟨 | 🟨 | 🟨 | | 🟥 |
| **UKI-SOUTHGRID-SUSX** | 🟩 | 🟩 | 🟩 | 🟨 | 🟩 | 🟩 | 🟨 | 🟨 | 🟩 | |

https://ps-dash.dev.ja.net/maddash-webui/index.cgi?dashboard=UK%20Mesh%20Config

Example perfSONAR results: Durham to Cambridge

IPv4 throughput

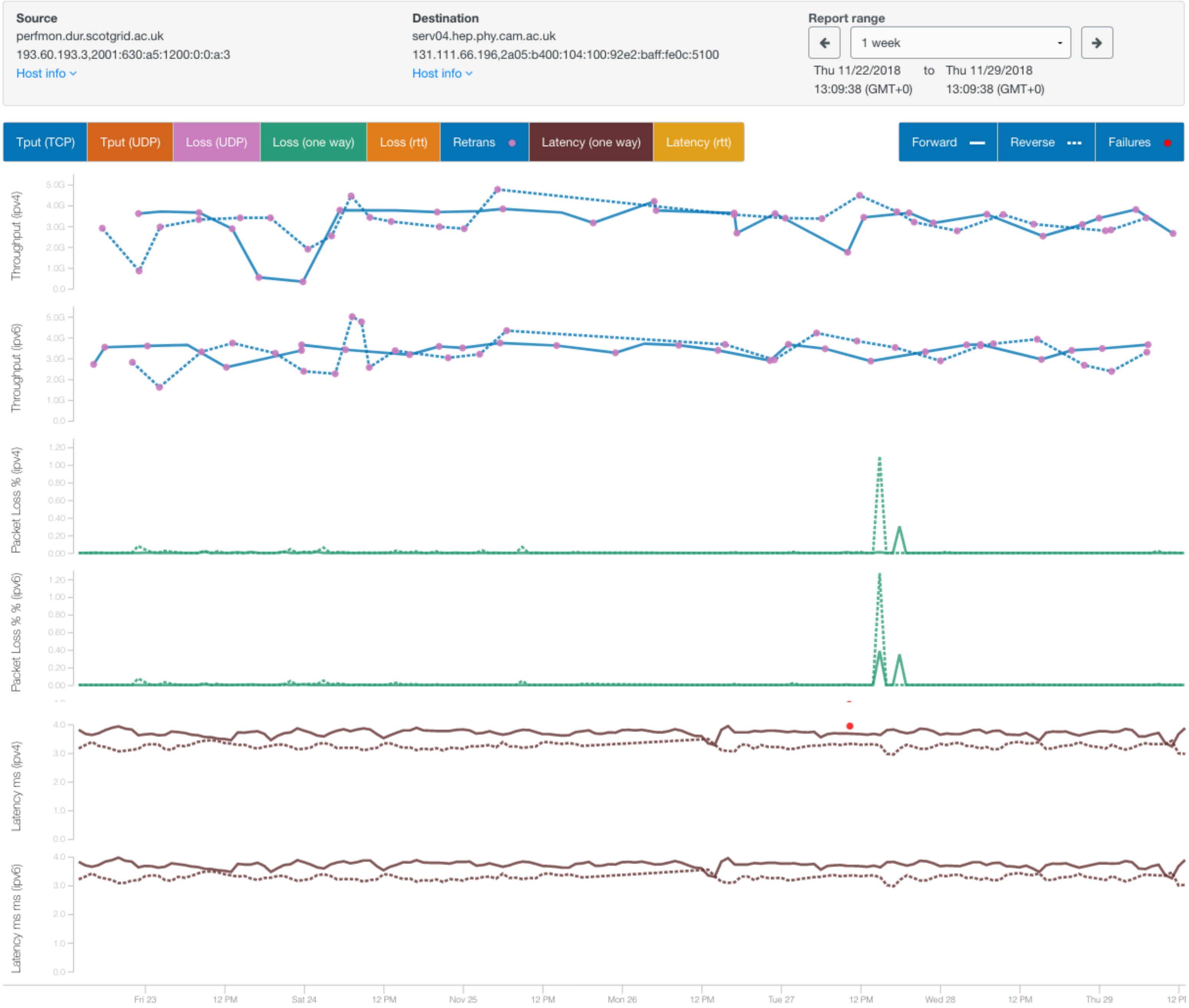
IPv6 throughput

IPv4 packet loss

IPv6 packet loss

IPv4 latency

IPv4 latency



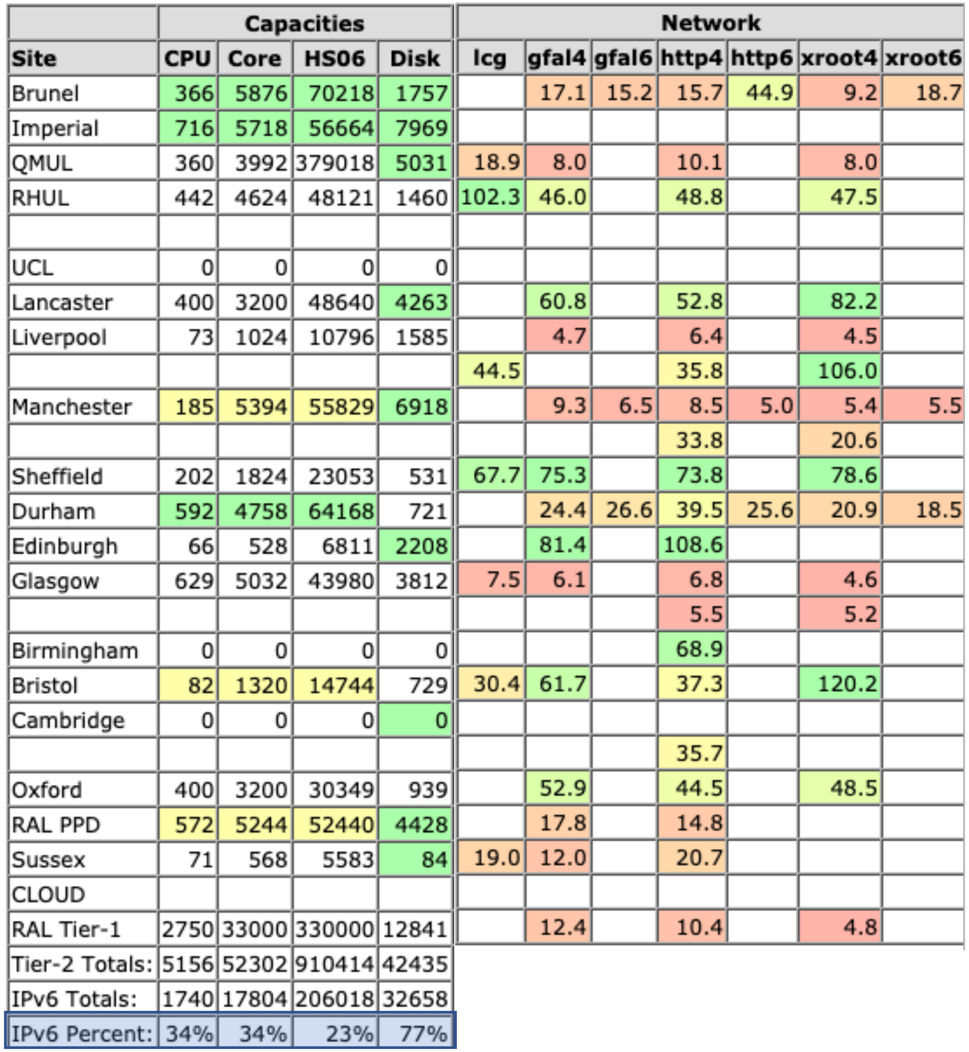
**Source**
perfmon.dur.scotgrid.ac.uk
193.60.193.3,2001:630:a5:1200:0:0:a:3
Host info

**Destination**
serv04.hep.phy.cam.ac.uk
131.111.66.196,2a05:b400:104:100:92e2:baff:fe0c:5100
Host info

**Report range**
1 week
Thu 11/22/2018 13:09:38 (GMT+0) to Thu 11/29/2018 13:09:38 (GMT+0)

# GridPP Network Tests

- Jobs are sent to a WN at each site to read 1GB, 2GB and 3GB files from each site's Storage Element (SE) using various protocols. The files have been previously replicated to all SE. The table shows average bandwidth (in MB/s) into the worker nodes computed from the times taken for each combination (including the local SE).
- Test over IPv6 also
- Transfers are made with lcg-cp, gfal-copy, curl and xrdcp over IPv4 and IPv6 (where relevant)
- Also recording the percentage of UK CPU and storage available over IPv6
- UK currently has 77% of Tier-2 disk storage available over IPv6



| | Capacities | | | | Network | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Site | CPU | Core | HS06 | Disk | lcg | gfal4 | gfal6 | http4 | http6 | xroot4 | xroot6 |
| Brunel | 366 | 5876 | 70218 | 1757 | | 17.1 | 15.2 | 15.7 | 44.9 | 9.2 | 18.7 |
| Imperial | 716 | 5718 | 56664 | 7969 | | | | | | | |
| QMUL | 360 | 3992 | 379018 | 5031 | 18.9 | 8.0 | | 10.1 | | 8.0 | |
| RHUL | 442 | 4624 | 48121 | 1460 | 102.3 | 46.0 | | 48.8 | | 47.5 | |
| | | | | | | | | | | | |
| UCL | 0 | 0 | 0 | 0 | | | | | | | |
| Lancaster | 400 | 3200 | 48640 | 4263 | | 60.8 | | 52.8 | | 82.2 | |
| Liverpool | 73 | 1024 | 10796 | 1585 | | 4.7 | | 6.4 | | 4.5 | |
| | | | | | 44.5 | | | 35.8 | | 106.0 | |
| Manchester | 185 | 5394 | 55829 | 6918 | | 9.3 | 6.5 | 8.5 | 5.0 | 5.4 | 5.5 |
| | | | | | | | | 33.8 | | 20.6 | |
| Sheffield | 202 | 1824 | 23053 | 531 | 67.7 | 75.3 | | 73.8 | | 78.6 | |
| Durham | 592 | 4758 | 64168 | 721 | | 24.4 | 26.6 | 39.5 | 25.6 | 20.9 | 18.5 |
| Edinburgh | 66 | 528 | 6811 | 2208 | | 81.4 | | 108.6 | | | |
| Glasgow | 629 | 5032 | 43980 | 3812 | 7.5 | 6.1 | | 6.8 | | 4.6 | |
| | | | | | | | | 5.5 | | 5.2 | |
| Birmingham | 0 | 0 | 0 | 0 | | | | 68.9 | | | |
| Bristol | 82 | 1320 | 14744 | 729 | 30.4 | 61.7 | | 37.3 | | 120.2 | |
| Cambridge | 0 | 0 | 0 | 0 | | | | | | | |
| | | | | | | | | 35.7 | | | |
| Oxford | 400 | 3200 | 30349 | 939 | | 52.9 | | 44.5 | | 48.5 | |
| RAL PPD | 572 | 5244 | 52440 | 4428 | | 17.8 | | 14.8 | | | |
| Sussex | 71 | 568 | 5583 | 84 | 19.0 | 12.0 | | 20.7 | | | |
| CLOUD | | | | | | | | | | | |
| RAL Tier-1 | 2750 | 33000 | 330000 | 12841 | | 12.4 | | 10.4 | | 4.8 | |
| Tier-2 Totals: | 5156 | 52302 | 910414 | 42435 | | | | | | | |
| IPv6 Totals: | 1740 | 17804 | 206018 | 32658 | | | | | | | |
| IPv6 Percent: | 34% | 34% | 23% | 77% | | | | | | | |

https://pprc.qmul.ac.uk/~lloyd/gridpp/ukgrid.html

GridPP
UK Computing for Particle Physics

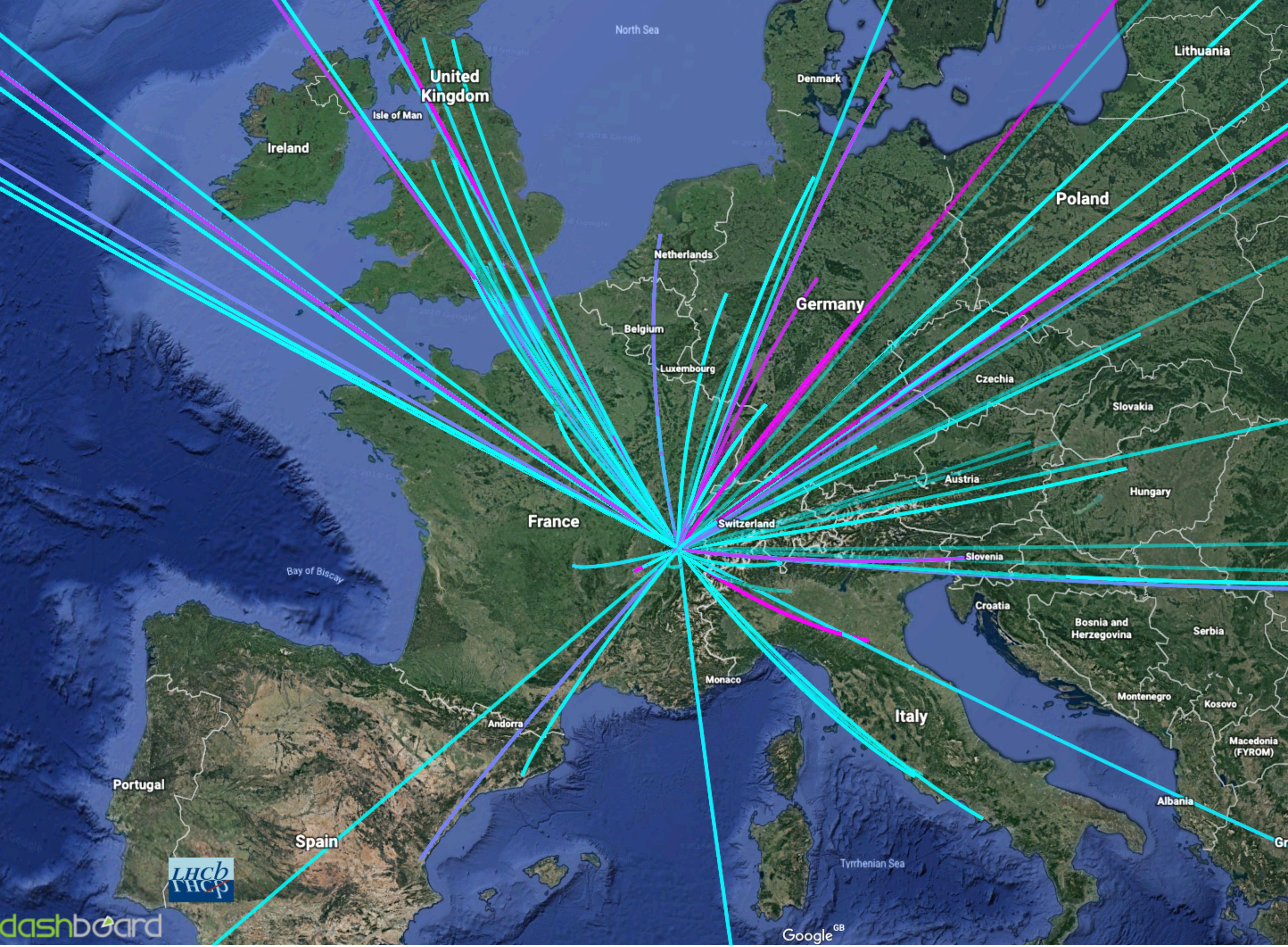
# Summary

- The WLCG needs to be ready for an offer of opportunistic IPv6-only CPU resources
- We are slowly but surely making our computing service IPv6 ready
- IPv6-only worker nodes at one WLCG site are already running production jobs
- Tier-1s and Tier-2s should be providing production storage accessible over IPv6 (93% and 64% are, respectively)
- This means 65% of LHC data is now accessible over IPv6
- The volume of data transferred over IPv6 has increased over the last year, 52% of bulk data transfers now go over IPv6
- 58% of WLCG perfSONAR hosts are now reporting 'IPv6-enabled'

- Finally, one hopefully positive, side-effect is that this is encouraging IPv6 adoption in a large number (~170) of research institutes worldwide

# Acknowledgements

- Reported work done by the HEPiX IPv6 Working Group and WLCG IPv6 Task Force and many others in the WLCG


North Sea
Lithuania
United Kingdom
Denmark
Isle of Man
Ireland
Poland
Netherlands
Germany
Belgium
Luxembourg
Czechia
Slovakia
Austria
Hungary
France
Switzerland
Slovenia
Croatia
Bosnia and Herzegovina
Serbia
Bay of Biscay
Monaco
Montenegro
Kosovo
Italy
Andorra
Macedonia (FYROM)
Portugal
Albania
Gr
Spain
Tyrrhenian Sea
LHCb
dashboard
Google GB