

**GridPP**

UK Computing for Particle Physics

GridPP Project Management Board

---

# LHC Network Forward Look 2024

---

Document identifier: **GridPP-PMB-207**

Date: **17 June 2024**

Version:

Document status: **Final**

Authors: **P.Clarke, T.Chown, D.Costanzo,  
A.Dewhurst, K.Ellis, D.Rand,  
A.Rogovskiy**

---

## Introduction

This document presents the UK networking capacity required by GridPP for LHC operations for the period 2024 - 2027. This is prepared at the beginning of 2024 during Run-3.

The scope of this report includes:

- The LHC schedule
- ATLAS, CMS and LHCb experiment requirements
- Scale of HL-LHC requirements and WLCG data challenges
- Tier-1 usage and connectivity
- Tier-2 connectivity
- IPv6 readiness
- LHCONE statement
- perfSONAR statement
- Jisc statement

## Executive Summary

The LHC is now in the middle of Run-3, which is due to finish at the end of 2025. The LHC has historically been very successful, leading to larger data rates/volumes than expected. The summary of the information from experiments and previous experience is:

- We expect that the CERN Tier-0 to RAL Tier-1 network requirement will rise to approximately 400 Gbit/s by 2027. The majority of Tier-0 to Tier-1 data will continue to be transferred via the LHCOPN for the foreseeable future. The Tier-1 joined the LHCONE in 2023 and this has allowed the Janet link to also be used to transfer data to and from CERN and the other Tier-1 sites if the capacity is required.
- The Tier-1 connection to Janet was upgraded to a 400 Gbit/s resilient link in mid 2024. We anticipate that this should be adequate for the remainder of Run-3 (it does not carry the primary CERN-RAL traffic).
- In the next 2 years, the “large” Tier-2 sites (Glasgow, Imperial, Lancaster, Manchester, QMUL, RAL) are likely to need up to 100 Gbit/s for all LHC traffic. In fact, the majority are already connected at 100 Gbit/s and the others are being upgraded, demonstrating excellent forward planning with Jisc.
- We expect the “other” Tier-2 sites to require 20 Gbit/s for LHC traffic, though there are some “medium” sites that currently have a high network I/O rate and a connection bandwidth similar to the larger sites.
- The summary of the scale of connection required is:
  - **Tier-1: 400 Gbit/s**
  - **Tier-2 (large): 100 Gbit/s**
  - **Tier-2 (other): 20 Gbit/s (and in some cases possibly 40 Gbit/s)**

- The RAL Tier-1 has joined LHCONe. The Imperial and RAL-Tier-2 sites will continue their LHCONe connectivity. GridPP now requests the remaining core sites to consider joining LHCONe if they can easily do so. Any non-core site wishing to join LHCONe is endorsed to do so.
- GridPP thanks Jisc for its continued pro-active engagement with the LHC programme which has led to the UK continuing to be in an excellent position in respect of network bandwidth planning.

## LHC Schedule

The LHC is currently in Run 3 of the LHC era which is scheduled to run through 2024 and 2025 and will finish at the end of 2025. The upgrade to HL-LHC starts with Long Shutdown 3 (LS3) at the beginning of 2026, scheduled to last for three years. The HL-LHC Run 4 is scheduled to commence at the start of 2029.

- 2024: Run-3; year 3. This year included a (network) data challenge (DC24 in Q1)
- 2025: Run-3; year 4
- 2026: First year of Long Shutdown 3. Data Challenge 26 will run this year.
- 2027: LS3; year 2

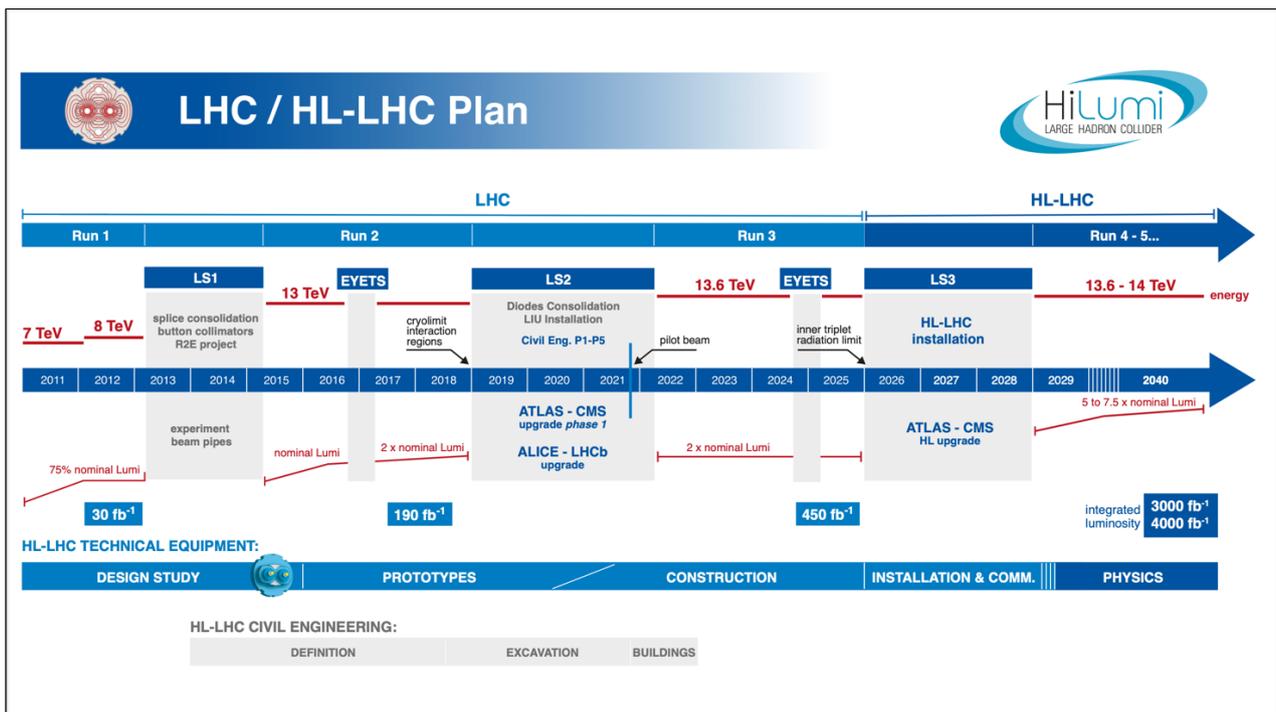


Figure 1: The LHC Schedule

## Experiment Requirements

Information from the individual experiments that sets the scale for network connectivity is provided below.

## ATLAS

During the remainder of LHC Run-3 and beginning of LS3 ATLAS expects a growth of processing and analysis capacity of a factor of  $\sim 2$  following the long-term trend of a 15-20% increase of resources per annum for both storage and CPU. These processes will either fetch data from storage and/or produce new datasets for storage. The architecture of the distributed system is changing, with an increased concentration of storage at fewer sites and the emergence of diskless sites with larger sites needing to serve remote CPUs. Consequently, there will be an additional increase in the bandwidth requirements. The overall computing model in terms of formats, versions and selections was revised for Run-3 with the creation of a new type of “derived AOD” (DAOD) dataset type, to be used by most ATLAS analysis groups. User analyses will access these DAODs directly over the network, which needs to be able to handle a fast analysis turnaround time.

This means that on the 4-year timescale, that the core (“nucleus”) ATLAS sites (QMUL, Manchester, Lancaster and Glasgow) will need at least  $\sim 80$  Gbit/s for the Tier-2 traffic alone (and hence a greater capacity at the campus site overall). It is possible that some of these sites may require 100 Gbit/s on this timescale. The UK needs to monitor the rate of growth of the network bandwidth at the larger ATLAS sites.

For the smaller non-core storage (“satellite”) sites, operational considerations require the site to have a demonstrated connection to a nucleus site of greater than 1.6 Gbit/s. In the UK, the satellite sites are actually relatively large and should be able to provide connections to  $\sim 4$  UK nucleus sites, with a weak scaling with the processing capacity available for ATLAS. Therefore, a bandwidth of 4-8 Gbit/s at the non-core sites should be sufficient with headroom for other traffic.

The summary for ATLAS expected by 2024 is shown in Table 1. The equivalent DC bandwidth is shown just to set a scale.

What	Volume moved in UK per-year (PB)	Equivalent DC bandwidth (Gbit/s)	Peak (if known) (Gbit/s)
Reconstruction	10	2.5	5
Derivation	100	25	50
Monte Carlo	56	14	28
Analysis	200	50	100
<b>Total</b>	<b>366</b>	<b>88</b>	<b>182</b>
<b>Tier-1 requirement</b>		<b>80</b>	
<b>Tier-2 (core) requirement</b>		<b>80</b>	
<b>Tier-2 (non-core) requirement</b>		<b>4 – 8</b>	

Table 1: ATLAS input figures. The final rows are the overall estimate of scale for Tier sites.

## CMS

The current network provision is sufficient for CMS at its Tier-1 and Tier-2 sites in 2023, and requirements are not expected to change significantly during the remainder of Run 3. However, we should note that in both 2022 and 2023 the amount of experimental data produced was below expectations and we anticipate significantly more in 2024. These data would arrive mostly at RAL Tier-1 disk and tape endpoints, before being later transferred to Tier-2 sites during ‘reprocessing’ campaigns.

CMS transfer throughput to the UK in the last year was dominated by jobs reading data directly from CERN, mostly to RAL Tier-1 and using the AAA system. Not only do CMS stream large volumes of ‘premix’ libraries

from CERN to sites running production jobs, but recently Tier-0 jobs were run at Tier-1 sites including RAL with the input data based at CERN.

One observation from the Tier-1 is that when the site has been in drain, and then returns to a full contingent of jobs, there is a period around two days later when the demands on network are particularly high. It is thought that this coincides with the time these jobs start the sub-job which reads the premix libraries. However, it is noted that since Tier-1 batch farm was connected to the LHCOPN, the performance of these jobs has improved significantly.

Improved XRootD monitoring should soon hopefully give us a much more accurate idea of how much data is being streamed via AAA, worldwide. Although people often like to say that CMS is using more network bandwidth for this method, the data has to get to the site running the jobs somehow, whether streamed via AAA or in a TPC transfer via FTS. Jobs using AAA only read the parts of the file needed.

So-called CMS ‘Tier-3’ sites are those without pledged resources to CMS, but these remain a good source of opportunistic CPU. Since these sites do not store CMS data, they use AAA for any input data required, and similarly return their outputs to native CMS sites within the UK. However, it is possible that some of these sites may suffer performance issues when reading remotely – especially those that are not located nearby to native CMS sites. This possibility warrants further investigation in the future.

Sites participated in the February 2024 data challenge, which highlighted the network needs looking forward to HL-LHC. The rates shown in Table 2 were requested from UK sites:

Site	Ingress (Gbit/s)	Egress (Gbit/s)
T2_UK_London_Brunel	3.8	0.7
T2_UK_London_IC	24.3	7.2
T2_UK_SGrid_Bristol	3.3	0.5
T2_UK_SGrid_RALPP	6.1	1.8

Table 2: Rates requested from UK sites in Data Challenge 24

These rates represent approximately 50% of a ‘best guess’ of network requirements for HL-LHC Run 4.

## LHCb

LHCb has recently finished the upgrade and commissioning of its detector. Year 2023 was mostly spent in commissioning, and not much real data were collected (data taking was also affected by several issues<sup>1,2</sup>). The year 2024 is effectively the first “full” data taking year.

According to the LHCb computing model<sup>3</sup>, the main source of external network traffic for the experiment is Tier-0 export, i.e. copying data from Tier-0 to Tier-1 sites, where the data is going to be processed. After the detector upgrade, the expected data rate from the HLT trigger is 10 Gbit/s. These data are then distributed to Tier-1 centers for further processing, leading to a minimal cumulative throughput for Tier-1 sites of 10 Gbit/s. The RAL Tier-1 is the biggest Tier-1 center for LHCb, accumulating about 30% of experiment storage

<sup>1</sup><https://indico.cern.ch/event/1261512/contributions/5298521/attachments/2608855/4506860/MPP%20-%20TE-VSC%20-%20VELO%20RF%20Foil%20Incident.pdf>

<sup>2</sup><https://home.cern/news/news/accelerators/lhc-leak-repair-short-photostory>

<sup>3</sup><https://cds.cern.ch/record/2319756/files/LHCb-TDR-018.pdf>

resources. The required network rate for RAL is 25 Gbit/s in the minimal scenario. This requirement is believed to stay valid until the end of Run 4.

Tier-2D (a Tier-2D site means one holding significant data) sites do not participate in Tier-0 export, and therefore require less network bandwidth. Tier-2D sites do receive stripping output data, but the rate is much smaller than Tier-0 export (the amount of stripped data should be smaller than the raw data by more than a factor of 2<sup>1</sup>, and most of it is stored in Tier-1 sites). It is difficult to estimate exact rates or volume of data since the simulation software for Run 3 data is still being commissioned, and there have been no “full” data taking years so far. However, if we assume that half of the cumulative Tier-2D data is replaced each year, then the total amount of data that is going to be transferred is approximately 1PB.

What	Volume moved in UK per-year (PB)	Equivalent DC bandwidth (Gbit/s)
Data	45	34
Monte Carlo	0.5	0.2
Analysis	0.5	0.2
<b>Tier-1 requirement</b>		<b>34</b>
<b>Tier-2 requirement</b>		<b>1 – 4</b>

Table 3: LHCb input figures. The final rows are the overall estimate of scale for Tier sites.

## Scale of HL-LHC requirements and WLCG data challenges

It is envisaged that the data rates coming out of the HL-LHC at commencement in 2029 will be significant<sup>4</sup>. To prepare for this the WLCG has initiated a series of data challenges (DC) running roughly every two years at increasing proportions of the 2029 rates. The first DC ran in 2021<sup>5</sup> with a target of 10% of the expected HL-LHC rate and in early 2024 a second challenge<sup>6</sup> was run at 25% of the HL-LHC rate. In order to plan for DC24 the LHC experiments calculated the ingress and egress rates for Tier-1 and Tier-2 sites for the various data flows<sup>7</sup> and these are shown for the UK sites in Table 4 together with extrapolations for the upcoming data challenges in 2026 and 2028. The rate displayed for the Tier-1 is the threshold goal set for DC24.

Site	DC24 (25%)		DC26 (50%)		DC28 (100%)	
	Ingress	Egress	Ingress	Egress	Ingress	Egress
RAL-LCG2	170	170	340	340	680	680
UKI-LT2-IC-HEP	22	7	44	14	87	29
UKI-LT2-QMUL	30	27	59	55	118	109
UKI-SCOTGRID-GLAS-GOW	13	12	25	24	51	47
UKI-NORTHGRID-LANCS-HEP	27	25	54	50	108	100
UKI-NORTHGRID-MAN-HEP	30	27	59	55	118	109
UKI-SOUTHGRID-RALPP	11	6	21	12	42	23

<sup>4</sup> <https://zenodo.org/records/5532452>

<sup>5</sup> <https://zenodo.org/records/5767913>

<sup>6</sup> <https://indico.cern.ch/event/1225415/contributions/5155042/attachments/2593516/4476291/Data%20Challenge%202024.pdf>

<sup>7</sup> <https://cernbox.cern.ch/s/w1kuXkjlY9fYZJF>

<b>UKI-LT2-Brunel</b>	3	1	7	1	13	3
<b>UKI-SOUTHGRID-BRIS-HEP</b>	3	1	6	1	11	2

Table 4: Site ingress and egress rates for data challenges (Gbit/s). Note, QMUL did not take part in DC24, but is a similar size to Manchester and thus the figures for Manchester have been used.

These extrapolations suggest that for the Tier-1 a connection of 400 Gbit/s and for the core Tier-2s a connection of 100 Gbit/s will be required for the next data challenge in 2026, but that for the 2028 challenge an upgrade to 800 Gbit/s is likely to be required for the Tier-1 and for some of the Tier-2s an increase to 200 Gbit/s will be required.

An issue of concern for Jisc during DC24 was the possibility of filling the 300G link between the GÉANT and Janet networks. This consists of two links - a primary connection *Lon* and a secondary reserve connection *Lon2*. These links carry general IP and LHCONE WLCG traffic to and from GridPP sites. Network traffic between CERN and the other Tier-1s to and from RAL usually travels over the dedicated Optical Private Network (LHCOPN). However, during the first part of the DC24 the OPN was inoperable due to a damaged cable under the North Sea with the consequence that traffic which usually flows over the OPN was rerouted onto the LHCONE and hence also used the GÉANT-Janet links. Our analysis suggests that the GÉANT to Jisc peering hit its capacity and remained fully loaded from 9pm GMT on 13th Feb until 10am GMT on 14th Feb when the general IP traffic was moved away to the GÉANT-Janet backup link, leaving the main link for just LHCONE. This means the network would have been a bottleneck for LHCONE and non-LHCONE traffic coming into Janet from GÉANT for those 13 hours. Interestingly, the link peaked at around 265G, not 300G. We believe this is caused by a configuration issue GÉANT-side that we saw previously in November 2022, when the (then) 2 x 100G GÉANT peering hit a cap of 165G rather than 200G. While one of the three 100G links comprising the aggregate did hit 100G, the other two between them could not exceed the 165G cap.

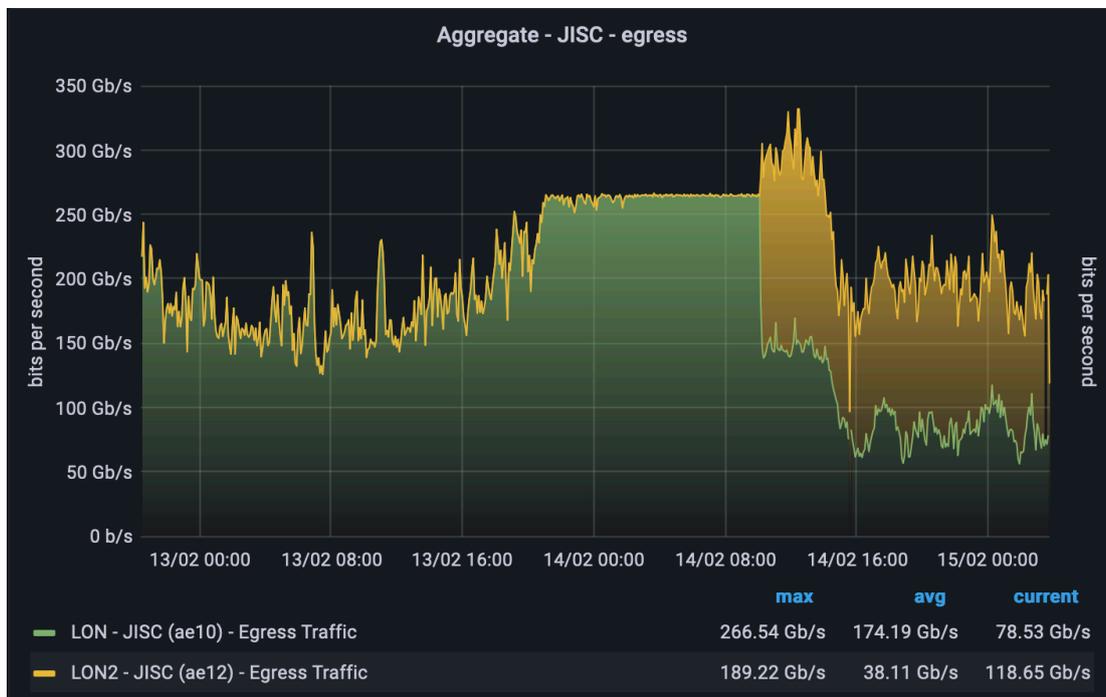


Figure 2 Network traffic between GÉANT and Janet during a part of DC24.

## TIER-1 usage and connectivity

At the end of 2023, the Tier-1 was finishing its migration to its new network. To reduce complexity during the migration external connectivity for the legacy and new networks was kept separate, i.e. they both had a separate link to JANET and were connected to one of the LHCOPN links. It was planned that in early 2024 the last remaining externally facing host on the legacy network would be migrated and both LHCOPN links would be attached to the new network. This greatly simplifies the logical setup and allows genuine full utilisation of the links. As this document is a forward look, for the remainder of the document we will assume that this consolidation has happened and that there is only one network at the Tier-1.

The Tier-1 network follows a spine/leaf design, with each leaf switch connected by 4 x 100 Gbit/s links to the spine. Storage has non-blocking network connectivity while CPU has a 3:1 overprovision. The Tier-1 network is connected to the rest of SCD (including Antares) via the Superspine with 16 x 100 Gbit/s links. External connectivity to the Tier-1 is provided by a resilient pair of leaf switches. These provide 2 x 100 Gbit/s LHCOPN links, 4 x 100 Gbit/s links to the site border routers for LHCONE traffic and 4 x 100 Gbit/s links to the site core for all other traffic.

From a routing perspective, external traffic is routed over the LHCOPN if possible, if not it will use the LHCONE and if that is not possible the default site link via the site core. Note that the LHCONE and default site link both use the same physical Janet link, however the LHCONE traffic bypasses the site firewall at RAL (and potentially at the other site as well).

Figures 3 and 4 show the external traffic on the LHCOPN and LHCONE respectively in 2023. Blue is traffic from RAL while green is traffic coming into RAL. The slightly different shades of each colour indicate different physical links. At the start of August the second 100 Gbit/s LHCOPN link was put into production and there is an immediate increase in the traffic. This wasn't an increase in total throughput as there is a corresponding decrease in LHCONE traffic. This is a result of the Worker Nodes being moved to the new network first when it was just connected to LHCONE but not LHCOPN. During that time if jobs (primarily CMS) downloaded data from CERN it used the LHCONE.



Figure 3 Traffic over the LHCOPN in 2023. Blue is traffic from RAL while green is traffic coming into RAL. The slightly different shades of each colour indicate different physical links. [The y scale runs from -30 to +70 Gbit/s, the x scale runs from week 2 to week 48]

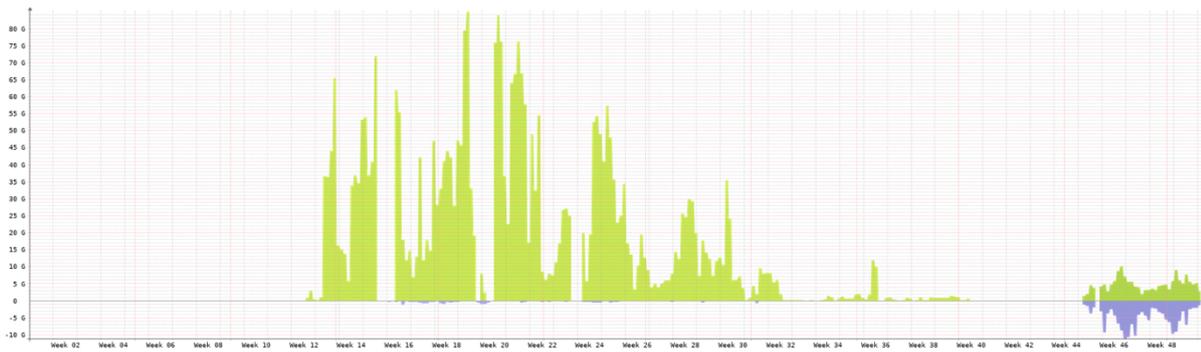


Figure 4 Traffic over the LHCONE in 2023. Blue is traffic from RAL while green is traffic coming into RAL. The slightly different shades of each colour indicate different physical links. [The y scale runs from -10 to +80 Gbit/s, the x scale runs from week 2 to week 48]

Figure 5 shows the traffic via the site core from the new network in 2023. The legacy network is not included as it has traffic from non-Tier-1 services. At the start of October Echo gateways were added to the new network and a significant increase in traffic is observed. This is primarily for FTS transfers to and from UK sites.



Figure 5 Non-LHCOPN or LHCONE traffic for the new network in 2023. Blue is traffic from RAL while green is traffic coming into RAL. The slightly different shades of each colour indicate different physical links. [The y scale runs from -4 to +5.5 Gbit/s, the x scale runs from week 2 to week 48]

By 2025 the following improvements are expected to be made:

- The RAL site link to JANET will be upgraded from 2 x 100 Gbit/s to 4 x 100 Gbit/s (resilient) in the first quarter of 2024 (now completed).
- The Tier-1 batch farm will be made dual stack by end of Q2 2024 (completed).
- The EOS frontend to Antares will be made dual stack and put on both the LHCOPN and LHCONE.
- Internal traffic will be directed over the Superspine and through a common SCD exit router.

The long-term planning continues to follow the plan set out in the RAL 2021 CHEP paper<sup>8</sup>. In the second half of GridPP7 (2026, 2027) the following improvements are expected to be made:

- The current Tier-1 spine will be replaced after 5 - 6 years in production with 400 Gbit/s capable switches.
- We will upgrade the LHCOPN link to 400 Gbit/s once this falls to the same level as the 2 x 100 Gbit/s links.

The RAL site link will be monitored and upgraded when necessary (up to 800 Gbit/s). At current growth rates this is expected to be in 2027 - 2028.

<sup>8</sup> [https://www.epj-conferences.org/articles/epjconf/pdf/2021/05/epjconf\\_chep2021\\_02074.pdf](https://www.epj-conferences.org/articles/epjconf/pdf/2021/05/epjconf_chep2021_02074.pdf)

## Tier-2 connectivity

Table 5 below summarises, for each Tier-2, the Institute connection to Janet, and the Tier-2 site connection itself. Comments have been solicited from all site admins and are also shown in the table. Current Tier-2 connections range between 10-100 Gbit/s.

Site	Institute Janet connection	Tier-2 site WAN connection	Forward Look	Comment from site administrator
UKI-LT2-IC-HEP	100 Gbit/s	100 Gbit/s shared		
UKI-LT2-QMUL	2*100 Gbit/s	100 Gbit/s dedicated		Site uses backup connection. Site capable of higher rates if needed. LHCONE also an option.
UKI-NORTHGRID-LANCS-HEP	40 Gbit/s	40 Gbit/s "dedicated"	100 Gbit/s "dedicated" in 2024	Dedicated in this case means using principle user of the "backup link". Cluster NAT'd
UKI-NORTHGRID-MAN-HEP	100 Gbit/s	40 Gbit/s dedicated		
UKI-SCOTGRID-GLASGOW	2x20 Gbit/s?	"2 x 20 Gbit/s"		Upgrade in progress (should be 4x20Gbit/s?? for INST?). Technically we're not supposed to use more than half the Inst link, but we definitely have gone above 20Gbits before...
UKI-LT2-Brunel	4 x 10 Gbps	4x10 Gbps		
UKI-LT2-RHUL	2 x 10 Gbit/s	10 Gbit/s dedicated	Upgrade to 2 x 10 Gbit/s or shared 100 Gbit/s, timescale uncertain.	
UKI-NORTHGRID-LIV-HEP	2 x 20 Gbps	2 x 10Gbps	Institute upgrade to 2 x 40Gbps early 2024.	
UKI-NORTHGRID-SHEF-HEP	4*10 Gbit/s	10 Gbit/s	No changes expected	I believe this to be correct our 10 Gbit/s is only capable of operating at 10 Gbit/s but I believe has 2 routes out of the building both capable of the full speeds. Central don't let me have any visibility over this.
UKI-SCOTGRID-DURHAM	10 Gbit/s	4 x 10 Gbit/s	Future planning 100 Gbit/s	We may see further increases beyond the planned 10 Gbit/s as more HPC systems come online at DURHAM.

<b>UKI-SCOTGRID-ECDF</b>	2 x 20 Gbit/s	2 x 25 Gbit/s across campus to the 2 x 20 Janet connection	Maybe 2 x 40 Gbit/s if traffic warrants.	ECDF is part of a shared University cluster.
<b>UKI-SOUTHGRID-BHAM</b>		2 x 10 Gbit/s (though one is currently broken)	Looking to upgrade site to 25 Gbit/s over the next few years depending on hardware money	
<b>UKI-SOUTHGRID-BRISTOL</b>		3 x 10 Gbit/s	2 x 10 Gbit/s	Probably 100 Gbit/s
<b>UKI-SOUTHGRID-OX-HEP</b>	100 Gbit/s	2 x 10 Gbit/s	Currently no upgrade plans	
<b>UKI-SOUTHGRID-RALPPD</b>	2x100 Gbit/s	2x100 Gbit/s	No Plans	
<b>UKI-SOUTHGRID-SUSX</b>				

Table 5: Network connectivity to Janet and forward look from each Tier-2 site.

Some additional detail regarding the situation at three of the Core Tier-2 sites is included below.

Manchester's Tier 2 is on 40 Gb/s currently, which is sufficient to support the (pre- HL-LHC) experiments. The university itself is on 100 Gbit/s. We are talking to the UoM network team to insert in their 2030 plan the upgrade necessary to sustain the 100 (150?) Gbit/s rates necessary by then.

Glasgow currently has access to a 20 Gbit/s link dedicated to research traffic, and steps have been taken recently at the campus level to improve the resilience of this connection. A second 20 Gbit/s link is available and work should commence shortly to make use of this. This will require the introduction of selective routing to distribute traffic between the two links; overall bandwidth will be 40 Gbit/s, but it will not operate as a single, transparent 40 Gbit/s link. We will continue to explore options for further increases in bandwidth with our campus network team and with Janet – in the future, we expect to require network connectivity comparable to the other core Tier-2 sites @ Imperial, Manchester, Lancaster, QMUL, but must also be able to make use of this internally.

It often advantageous in a collaboration for one or more members to act as pathfinders, exploring future technologies or operating regimes. Imperial College has in the past acted in this capacity in the area of networking by, for example, having a large connection bandwidth to the Janet network, joining LHCONe, and in the early adoption of IPv6. At Imperial College whilst the existing connection to the Janet network is 100 Gbit/s, the internal network between the cluster and the Janet connection is already capable of operating at 200 Gbit/s. It would be useful for Imperial to receive an upgrade of the Janet connection from 100 Gbit/s to 200 Gbit/s in the near future, for testing purposes, both to confirm that the site internal networking and storage infrastructure can indeed make use of such bandwidth, but also to see what the extra capacity affords in terms of increased workflow flexibility and how this effects the experiment's use of the site.

## IPv6 readiness

Guided by the HEPiX IPv6 working group, the WLCG has been a leading adopter of IPv6 within the Research and Education community showing leadership for similar projects such as the Square Kilometre Array. An initial request by the WLCG Management Board in 2017 to make storage and perfSONAR hosts dual-stack has now been largely completed by GridPP sites. The WLCG Management Board recently requested that sites now make their worker nodes dual-stack. The current situation at GridPP sites may be seen in Table 6 and on the WLCG IPv6 Task Force wiki<sup>9</sup>. The focus of the WLCG IPv6 Task Force now is to identify remaining IPv4 use with the view to enable a cutover to IPv6-only, perhaps on the LHCOPN first.

Site	Storage IPv6 dual-stack (if applicable)	Compute IPv6 dual-stack	Comments
UKI-LT2-IC-HEP	Yes	Yes	Site is fully IPv6 dual stack
UKI-LT2-QMUL	Yes	Yes	Site is fully IPv6 dual stack
UKI-NORTHGRID-LANCS-HEP	Yes	No	Compute v6 poses challenges that need to be considered.
UKI-NORTHGRID-MAN-HEP			
UKI-SCOTGRID-GLASGOW	Yes	No	Intention is to move compute to v6 dual-stack within the next few months Some weirdness with v6 for us in general as well...
UKI-LT2-Brunel	Yes	Yes	Site has been fully dual stack for maybe a decade.
UKI-LT2-RHUL	N/A	No	Compute v6 is a major challenge for us as we currently rely on NAT as part of the security model. Local IT support is also likely to be inadequate.
UKI-NORTHGRID-LIV-HEP	Yes	No	Should be able to move compute to dual-stack, subject to available effort. Will be investigating next year.
UKI-NORTHGRID-SHEF-HEP	N/A	No	Working towards it, but still a number of months away from having any batch system dual stacked.
UKI-SCOTGRID-DURHAM	No	Yes	Storage to follow after current issues are rectified
UKI-SCOTGRID-ECDF	N/A	No	
UKI-SOUTHGRID-BHAM-	Yes	No	Compute should be ~trivial but need help from central IT and the person who does it is off sick

<sup>9</sup> <https://twiki.cern.ch/twiki/bin/view/LCG/Wlclpv6#IPv6Comp>

HEP			
UKI-SOUTHGRID-BRIS-HEP	Yes	Yes	Worker nodes cannot be on IPv6 as long as we have HDFS. Will change once on CephFS
UKI-SOUTHGRID-OX-HEP	N/A	Yes	CE and worker nodes are dual-stack.
UKI-SOUTHGRID-RALPPD	Yes	Yes	Site is fully IPv6 dual stack. Planning to test IPv6 only+XLAT+NAT64 WN
UKI-SOUTHGRID-SUSX			

Table 6: IPv6 status of each Tier-2 site.

## LHCONE statement

GridPP has historically had no bandwidth limitation issues that would, in themselves, motivate the use of LHCONE and this remains the case today. This is in part due to the foresight of our sites and in part due to the proactive vigilance of Jisc for which we thank them.

GridPP recently requested that its Tier-1 site should join LHCONE. This was mainly for uniformity with all other Tier-1s and the rest of the WLCG. This has now happened.

GridPP has hitherto not specifically called for its Tier-2 sites to join LHCONE. This was because there has been no operational imperative and the other possible motivations (security) have not been sufficient to outweigh the effort overhead at Tier-2 sites, with respect to overstretched network groups. However, two sites voluntarily joined LHCONE (Imperial College and RAL PPD) and they were endorsed to do so by GridPP.

We note that Jisc have always been ready and willing to configure LHCONE for GridPP if we had requested it.

The present status in 2024 is that there remains no pure operational requirement, based upon bandwidth alone, for Tier-2 sites to join LHCONE and this situation is unlikely to change. There are, however, other motivations, and these will become more and more of an imperative with time. These include an increasing security risk where Tier-2 sites are outside a site firewall and open to the world. While there are multiple ways of mitigating security risks, joining the LHCONE means that you can give more trusted access to a significant fraction of WLCG sites. Also, In the medium to long term we expect to see a significant growth in unscheduled network traffic between sites. For example, as compute resources become more diverse, we may well see jobs running at one site accessing data interactively at another. Allowing GridPP resources at one site full access to other sites could in the long term be beneficial. Once again, the LHCONE is not the only solution to this problem, but it is an established one.

Principally for reasons of uniformity with the rest of WLCG, since August 2023 GridPP has modified its policy for core Tier-2 sites. GridPP now requests core sites to consider joining LHCONE if they can easily do so. This implies that they want to, they are technically capable of doing so, and their network groups are amenable. It is important to state that there is no implied pressure to do so, and it is understood that there may

be many reasons that it is not timely for a given core site.

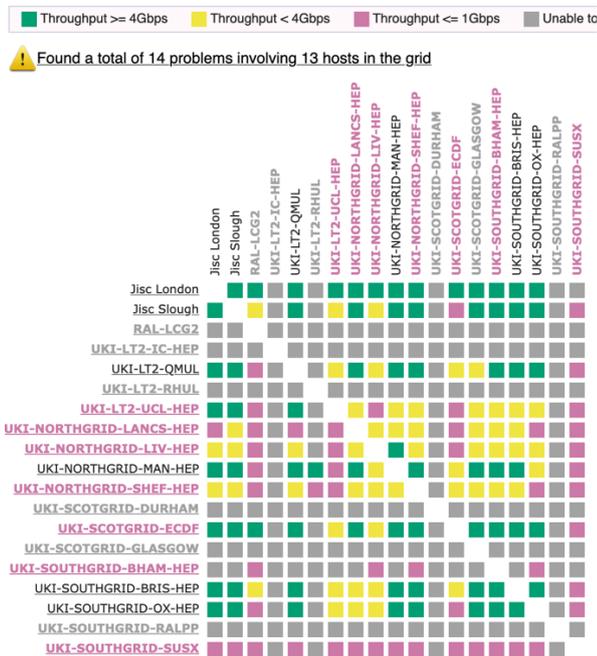
Notwithstanding the motivations given above, there is no change to the policy for non-core Tier-2 sites. GridPP does not specifically request its non-core sites to join LHCONE. Iteration with such sites had indicated that the “effort overhead argument” in general remains valid, with particular worry about burdening their network groups (although those sites which sit outside firewalls may consider whether the possible benefits outweigh such effort). As in the past, any non-core site wishing to join LHCONE is endorsed to do so.

For information we note that given that the RAL Tier-1 is now on LHCONE, then the routing to it from Tier-2 sites on LHCONE will bypass the firewall but the routing from those not on LHCONE will go through the RAL firewall. This is not expected to have any operational impact but may give an extra longer-term motivation for the core Tier-2 sites that may be exchanging data with the RAL Tier-1 to join LHCONE.

## PerfSONAR and Network Monitoring

PerfSONAR is a useful toolkit for the characterisation of network performance and the troubleshooting of network issues. The WLCG is probably the largest user community using perfSONAR and GridPP has been running perfSONAR for several years and continues to maintain and upgrade its perfSONAR infrastructure. The status of the IPv4 and IPv6 throughput meshes is shown in Figure 6. Janet Netsight<sup>10</sup> and BRIAN monitoring from GÉANT<sup>11</sup> monitoring may also be used to visualise the network.

UK Mesh Config - UK IPv4 Bandwidth - Throughput



UK Mesh Config - UK IPv6 Bandwidth - Throughput

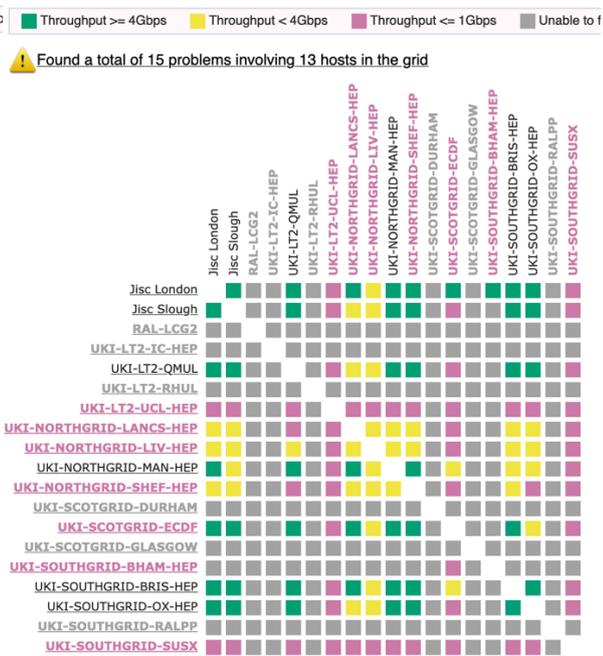


Figure 6 GridPP IPv4 and IPv6 perfSONAR throughput meshes.

<sup>10</sup> <https://netsight3.ja.net/view/>

<sup>11</sup> <https://public-brian.geant.org/d/S8qjUmwMz/jisc?orgId=5&from=now-24h&to=now>

## Jisc Statement

Jisc is in close communication with GridPP about its anticipated networking requirements and is taking these into account in the capacity planning of the Janet network. It is tracking the level of networking resources being discussed by the various communities that GridPP is involved with, most significantly, the WLCG at its LHCOPN/LHCONE meetings. Jisc is aware of the significant data rates expected with the advent of the High Luminosity upgrade to the LHC. It is also aware of the data challenges being planned by the WLCG leading up to the start of the HL-LHC in 2029, aimed at demonstrating the ability of the WLCG to transfer data at increasingly higher throughputs.

There is an ongoing programme of upgrading the connections of individual GridPP sites to the Janet network as and when requirements indicate it is appropriate. Also, following from this, upgrades to the Janet backbone, and the GÉANT peering, and we assist with the CERN-RAL OPN, and facilitate LHCONE implementation for two GridPP sites.

The Jisc Network Performance Team was set up to help Janet-connected sites make best use of the Janet Network, especially when transferring the large volumes of data often found in research computing. Particle physics, especially the WLCG, is one of these research communities and the Janet Netperf Team works closely with GridPP to the mutual benefit of both parties. Two members of GridPP are currently seconded part time to the Janet Netperf Team. The Jisc Netperf Team also works closely with the perfSONAR developers and the WLCG perfSONAR administrators and is involved in several WLCG network research and development activities in the areas of monitoring and measurement (e.g. packet marking), traffic engineering and the enhancement of throughput performance (e.g. jumbo frames and TCP-BBR).

## Summary of requirements

Whilst the nature of bandwidth estimation is imprecise, and depends heavily on the evolving computing models, the LHC schedule, and arbitrary headroom factors, the information provided above sets a scale for the aggregate bandwidth requirement by 2027 as:

**Tier-1: 400 Gbit/s**

**Tier-2 (large): 100 Gbit/s**

**Tier-2 (other): 20 Gbit/s (and in some cases 40 Gbit/s)**