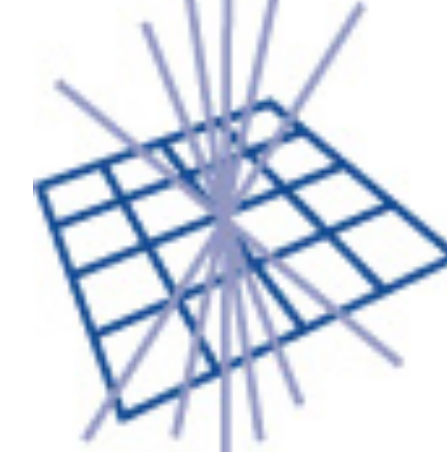




Science and
Technology
Facilities Council



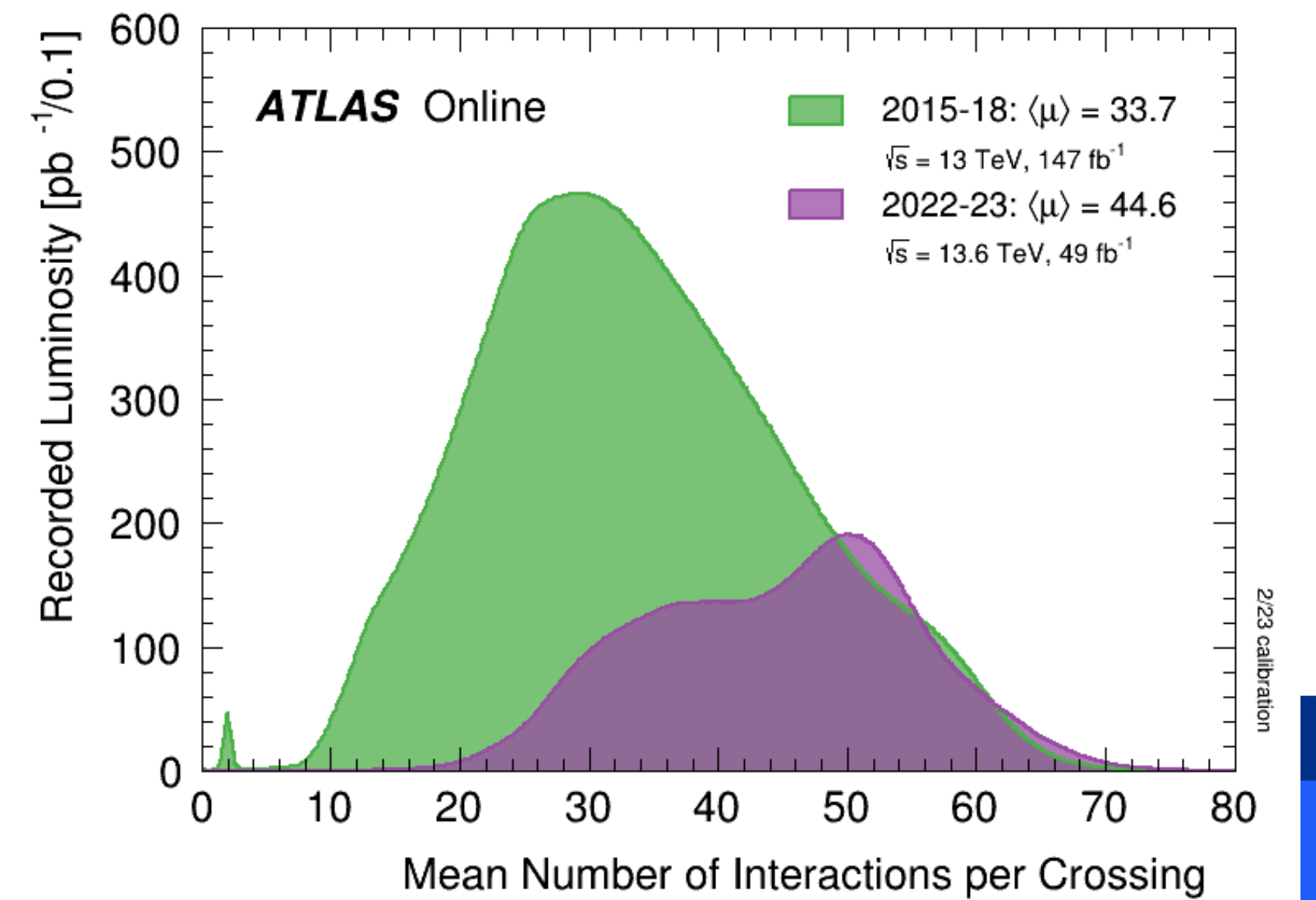
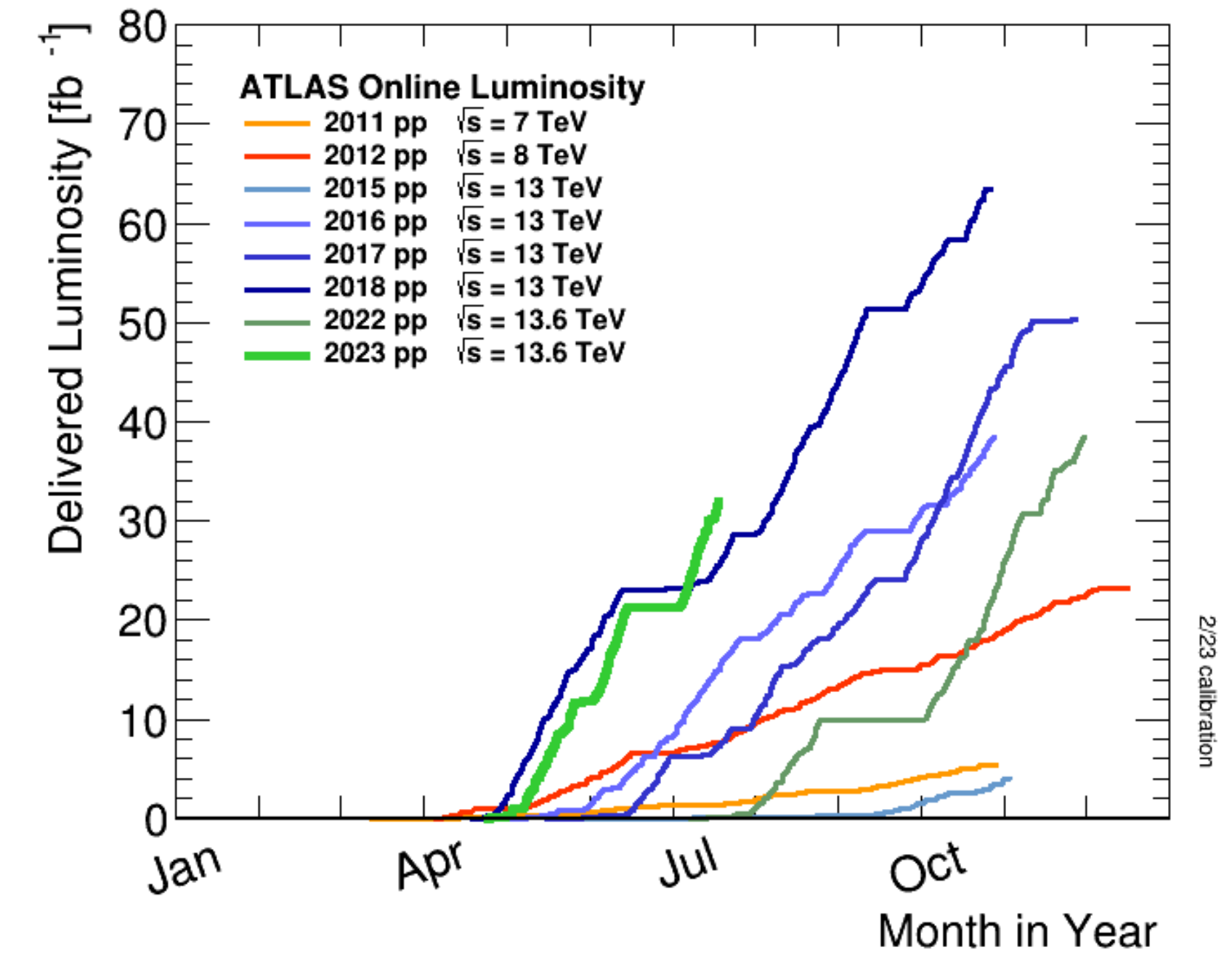
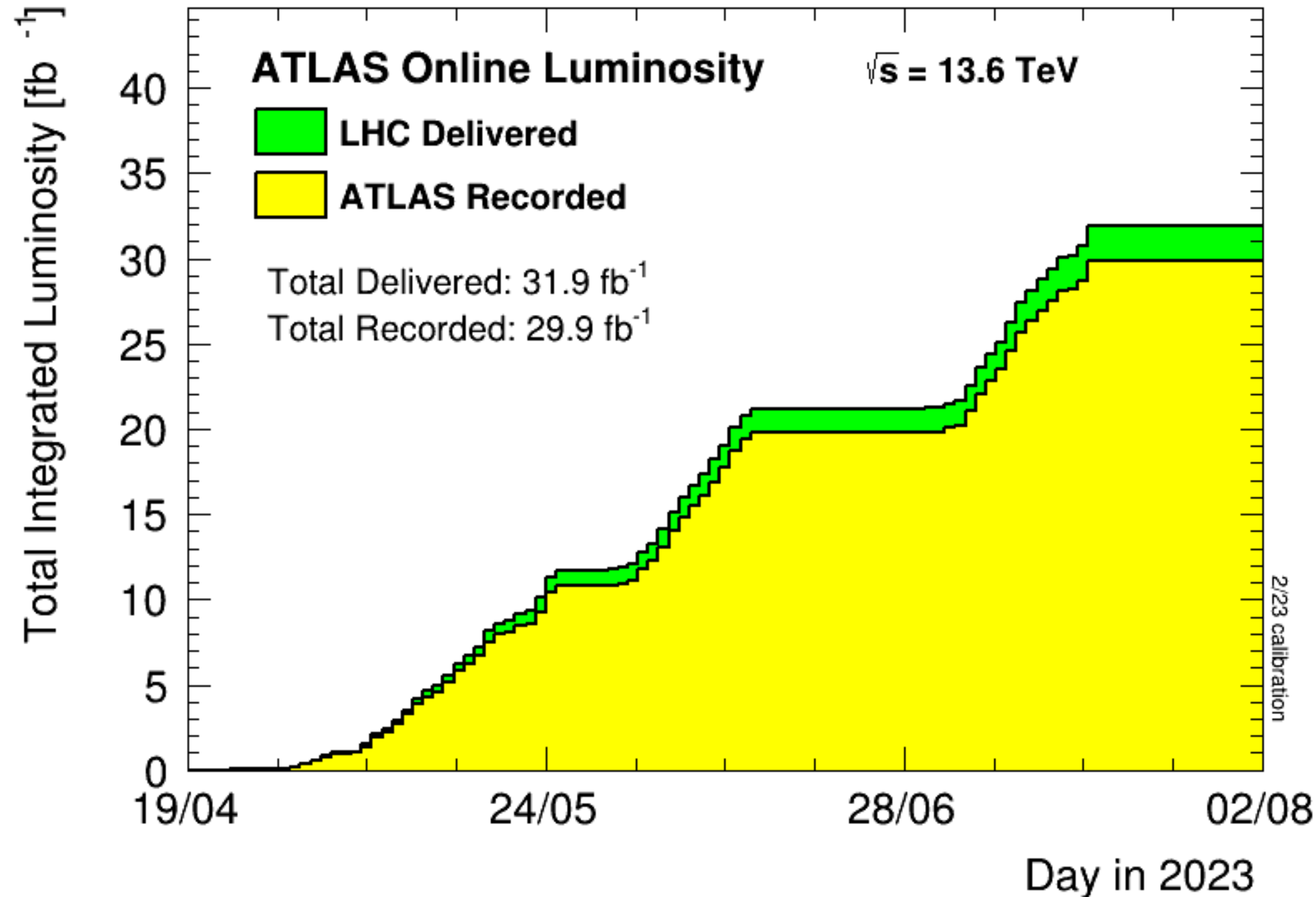
GridPP
UK Computing for Particle Physics

ATLAS @ GridPP 50

J. Walder on behalf of UK Cloud Squad
Ambleside, 1 Sept 2023

Status

- ~ 30fb⁻¹ recorded by ATLAS prior to problem on Monday, 17 July following LHC magnet quench.
 - Heavy ion running following some initial setup runs (~mid-Sept)

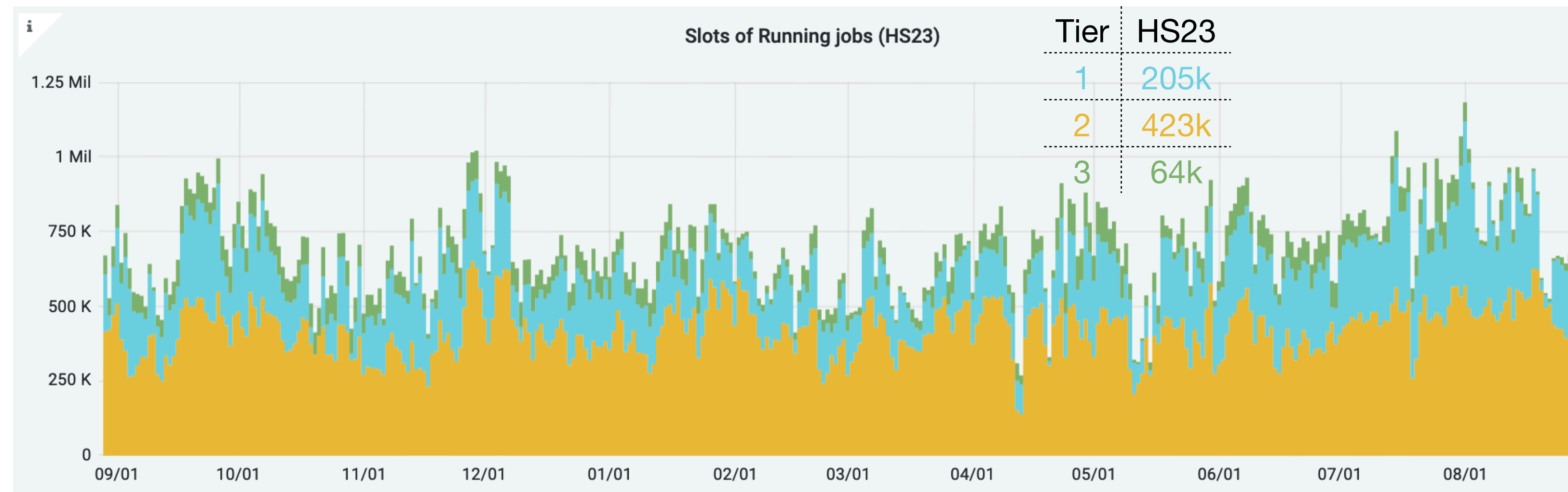


UK Compute (1)

- UK Delivered average of 692k HS23 over last year,

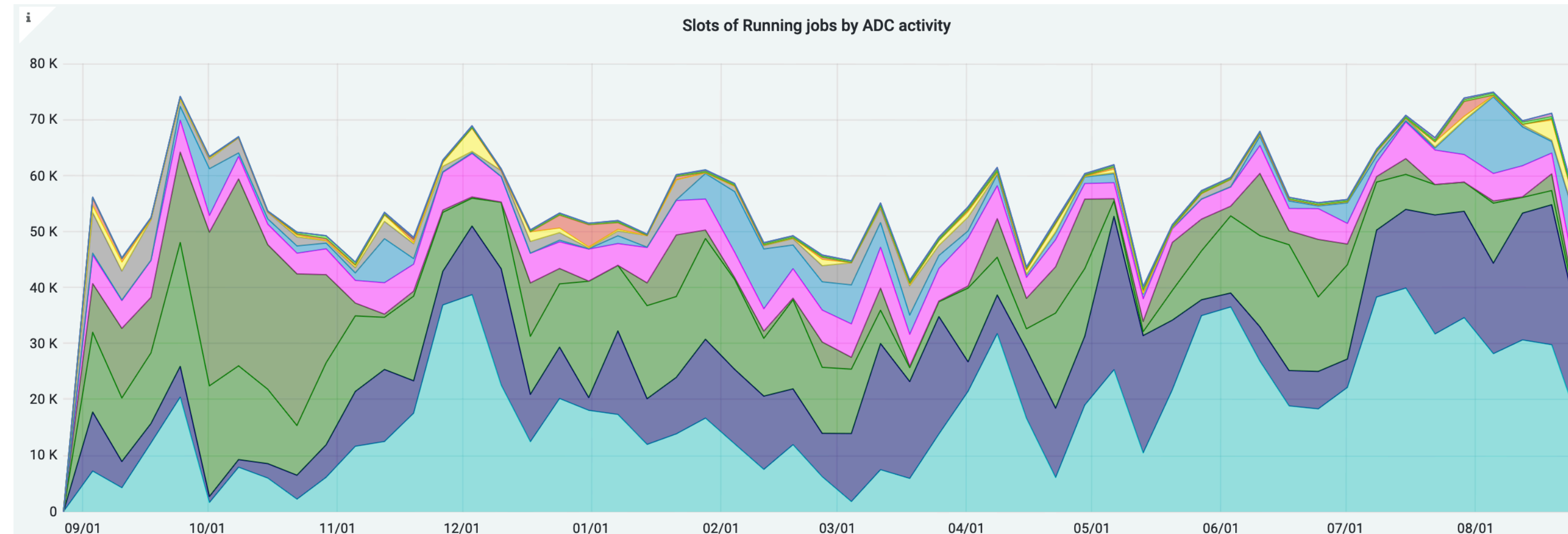
kHS23	2022	2023	Delivered
RAL	193	216	205
T2(+3)	261	297	487
Total	454	513	692

- Peaked at 1.3M HS23

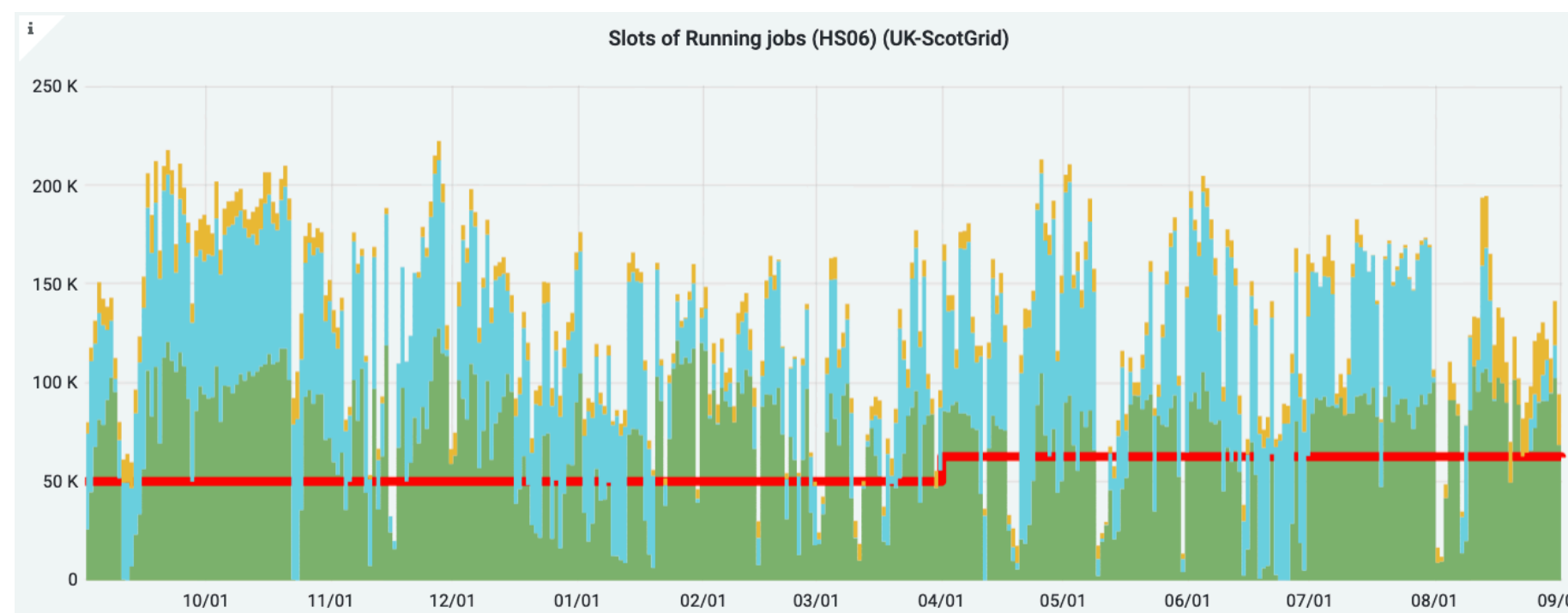
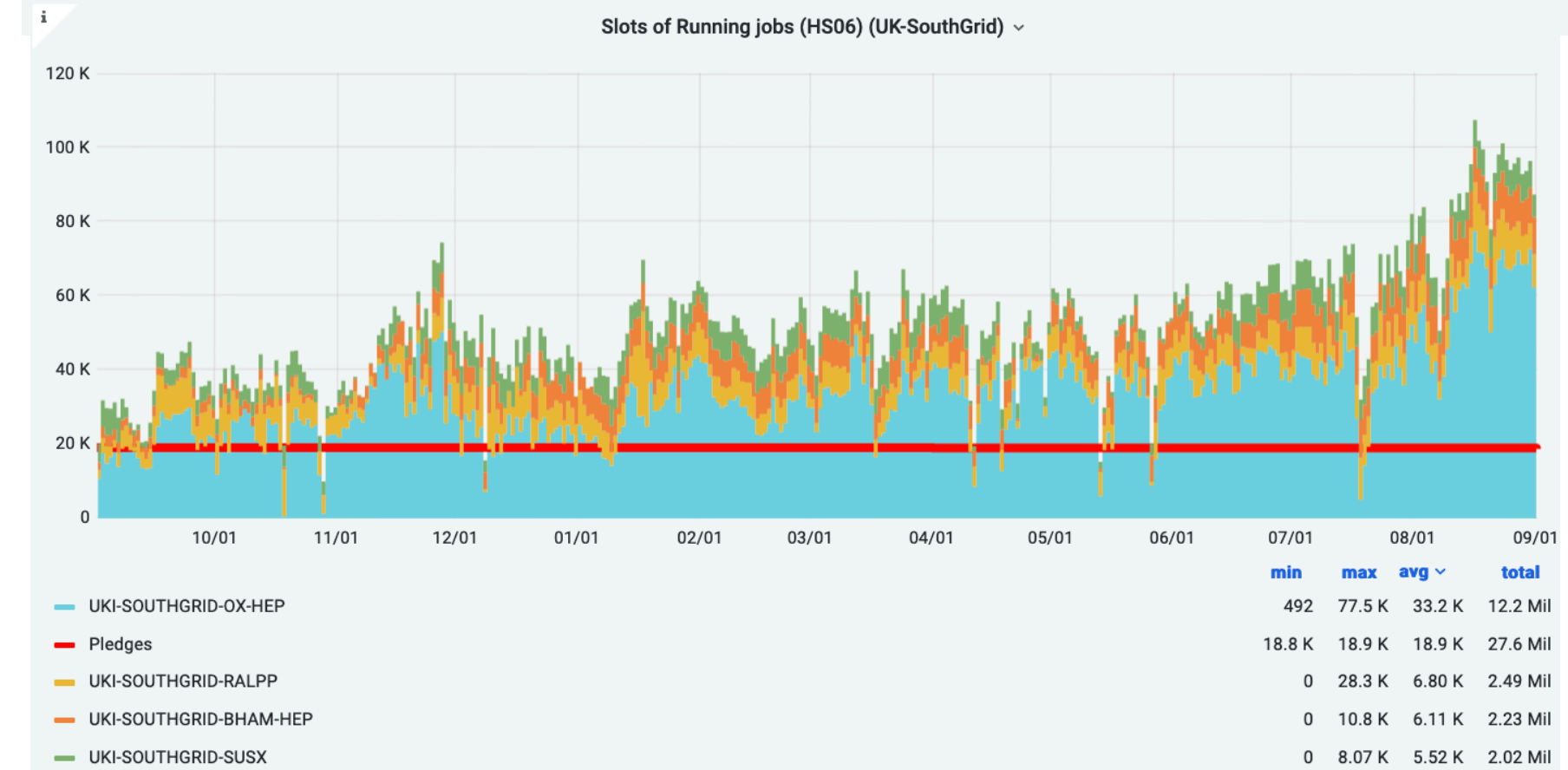
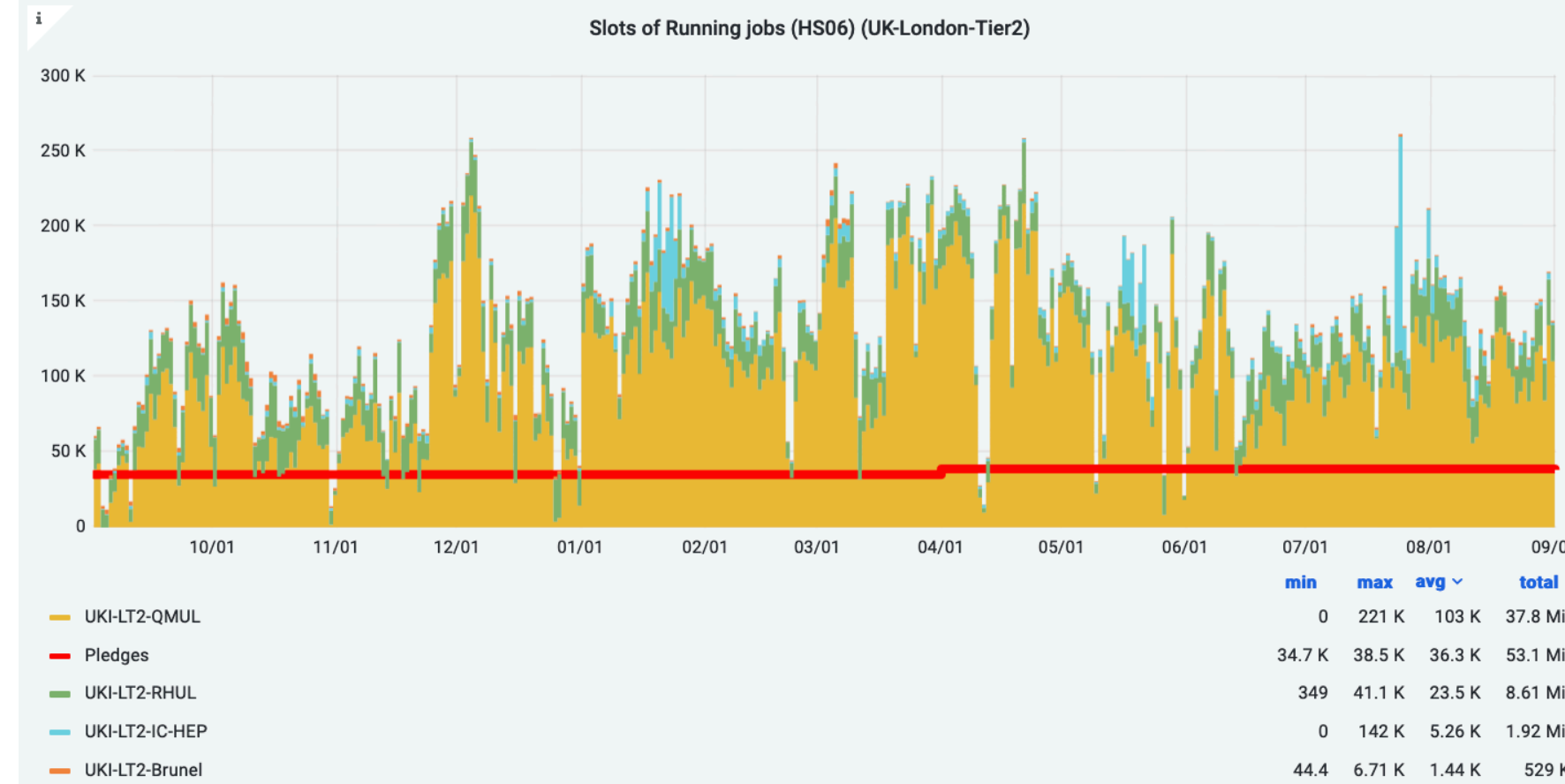
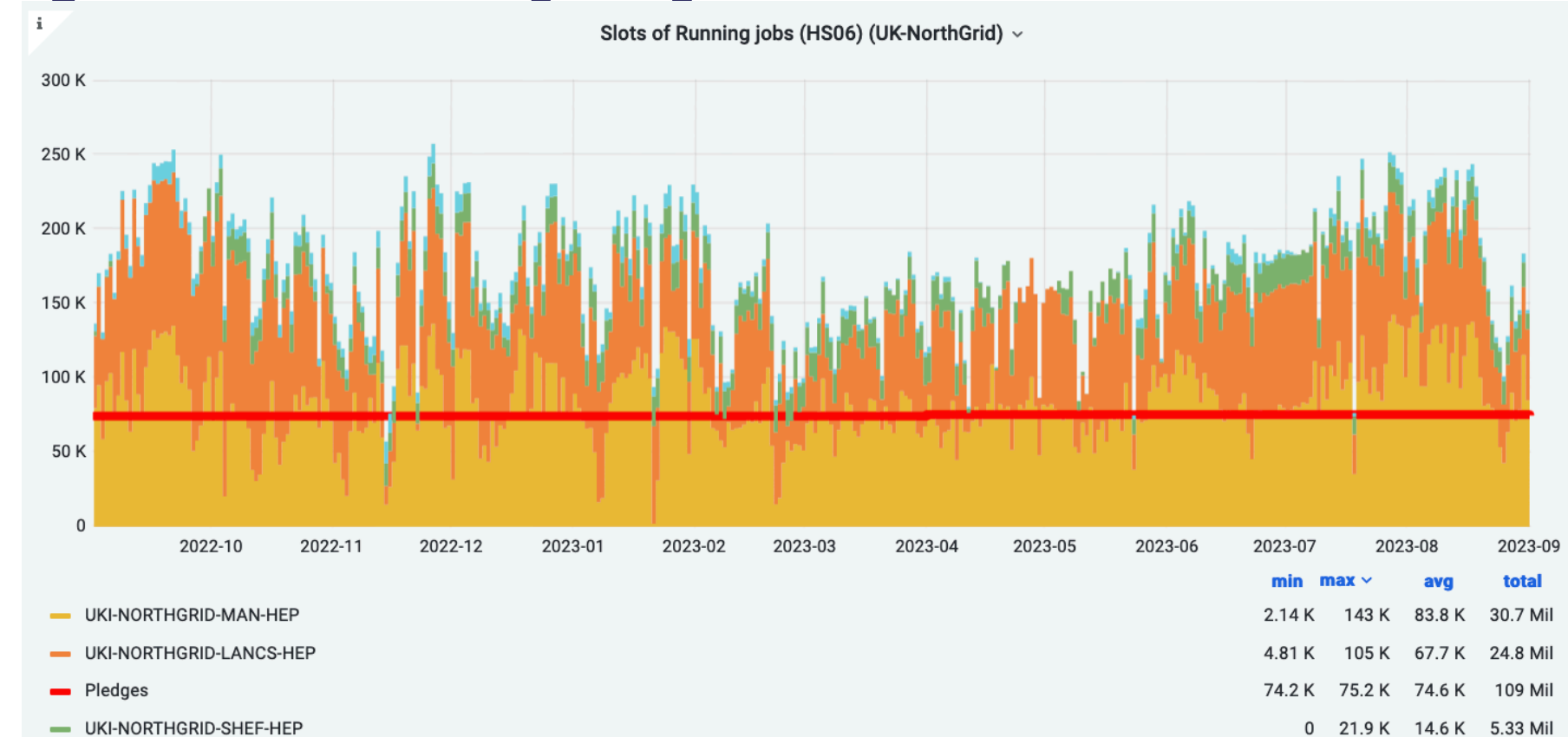
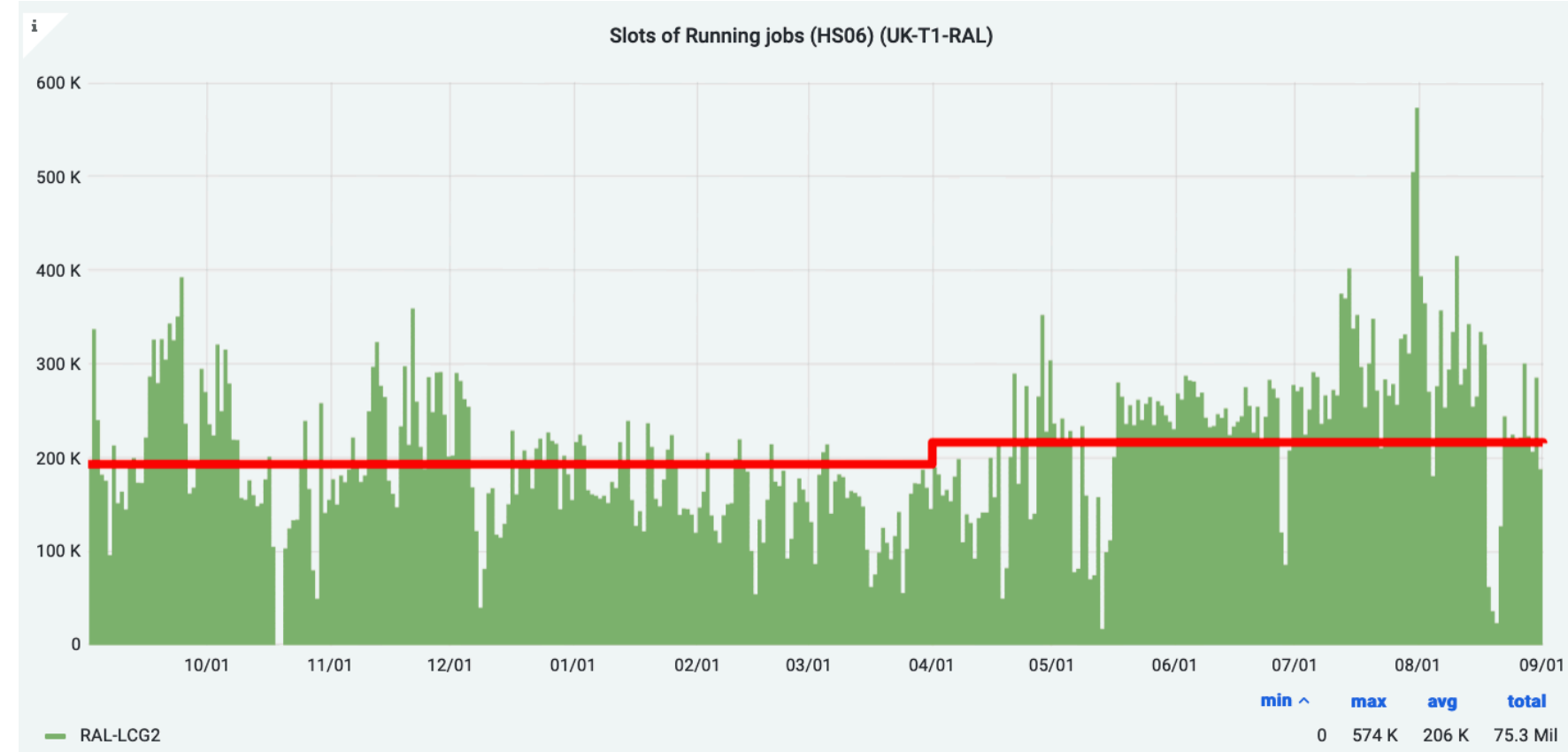


- CPU time spent mainly in MC related tasks

	max	avg
MC Simulation Full	39.9 K	17.8 K
MC Reconstruction	27.4 K	10.9 K
Group Production	22.5 K	10.8 K
MC Event Generation	33.4 K	5.84 K
User Analysis	8.58 K	4.83 K
MC Simulation Fast	13.7 K	2.61 K
Group Analysis	7.19 K	1.40 K
Data Processing	4.18 K	535
MC Resimulation	3.99 K	334



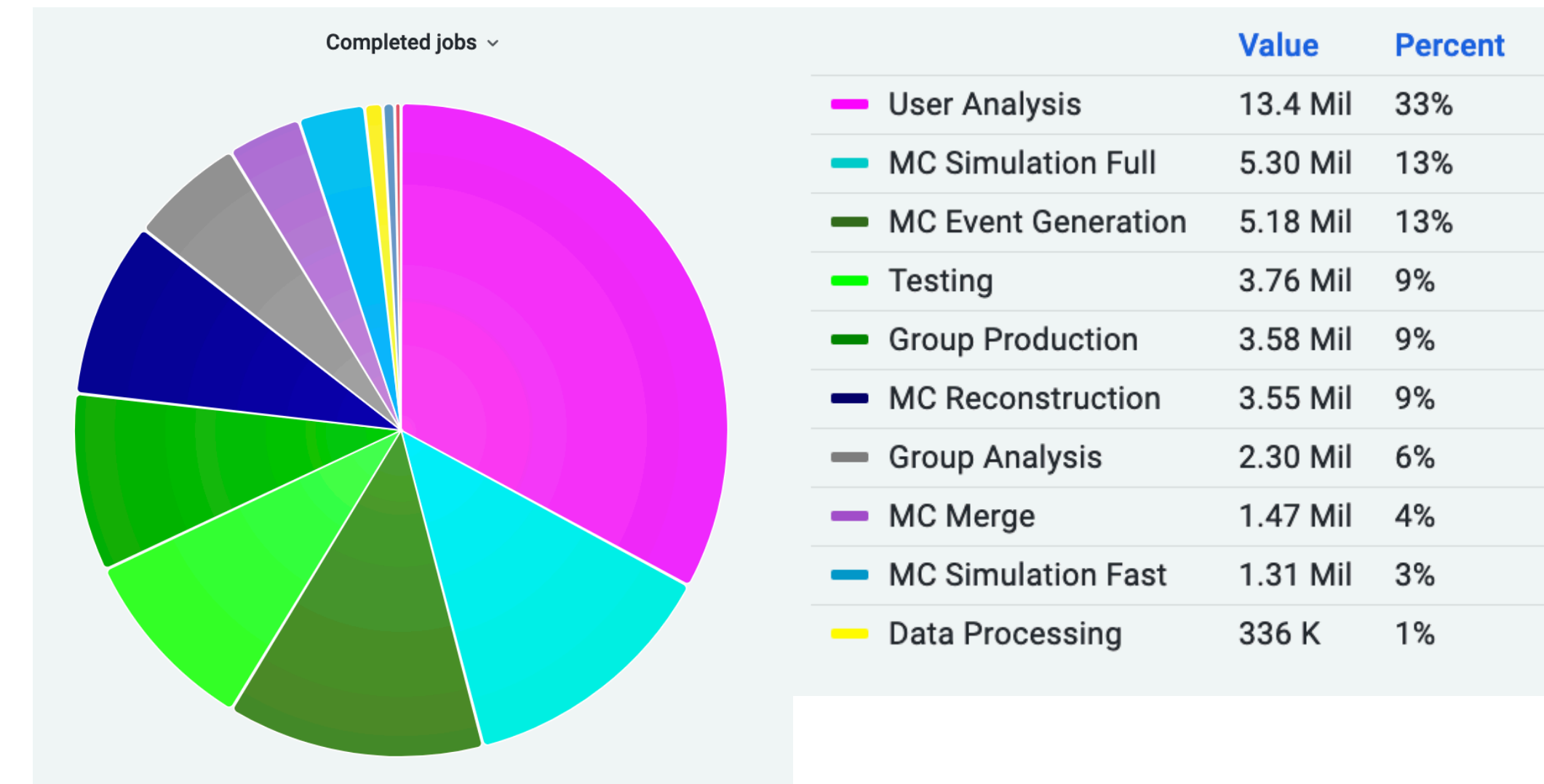
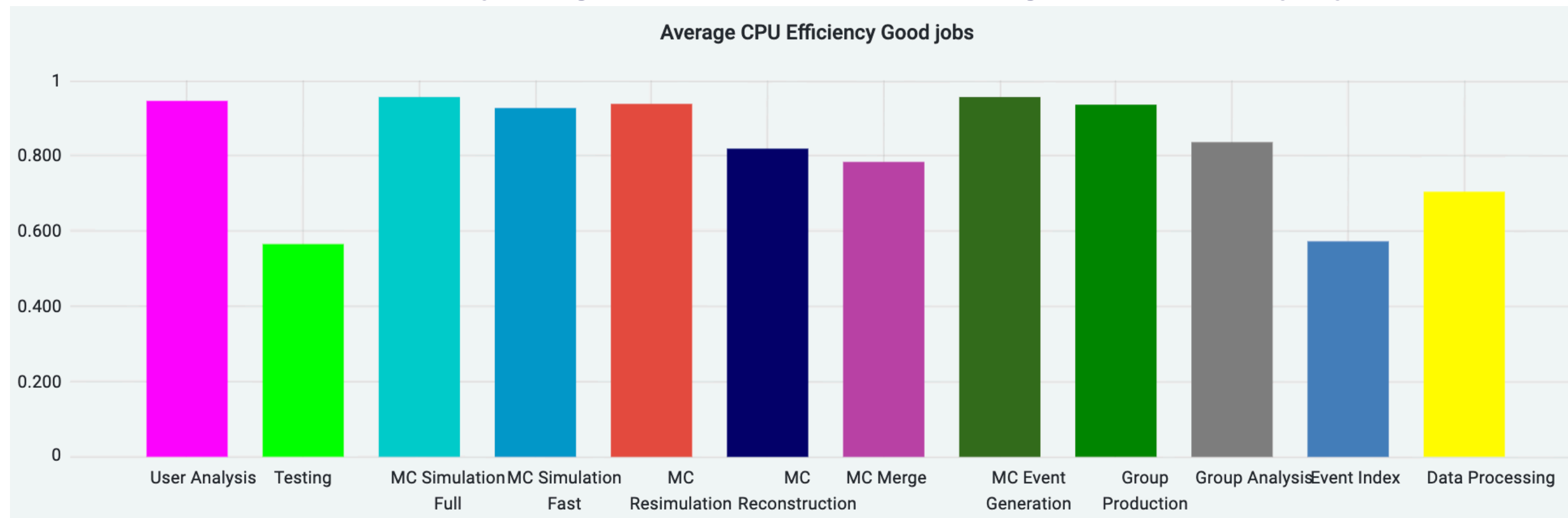
UK Compute (2)



	min	max	avg	total
UKI-SCOTGRID-GLASGOW	0	128 K	70.0 K	25.6 Mil
Pledges	49.9 K	62.5 K	55.2 K	80.7 Mil
UKI-SCOTGRID-DURHAM	0	109 K	51.3 K	18.8 Mil
UKI-SCOTGRID-ECDF	0	35.6 K	8.96 K	3.28 Mil

UK Compute (3)

- User analysis contributes ~ 1/3 of completed jobs
- CPU efficiency is good across the range of activity types



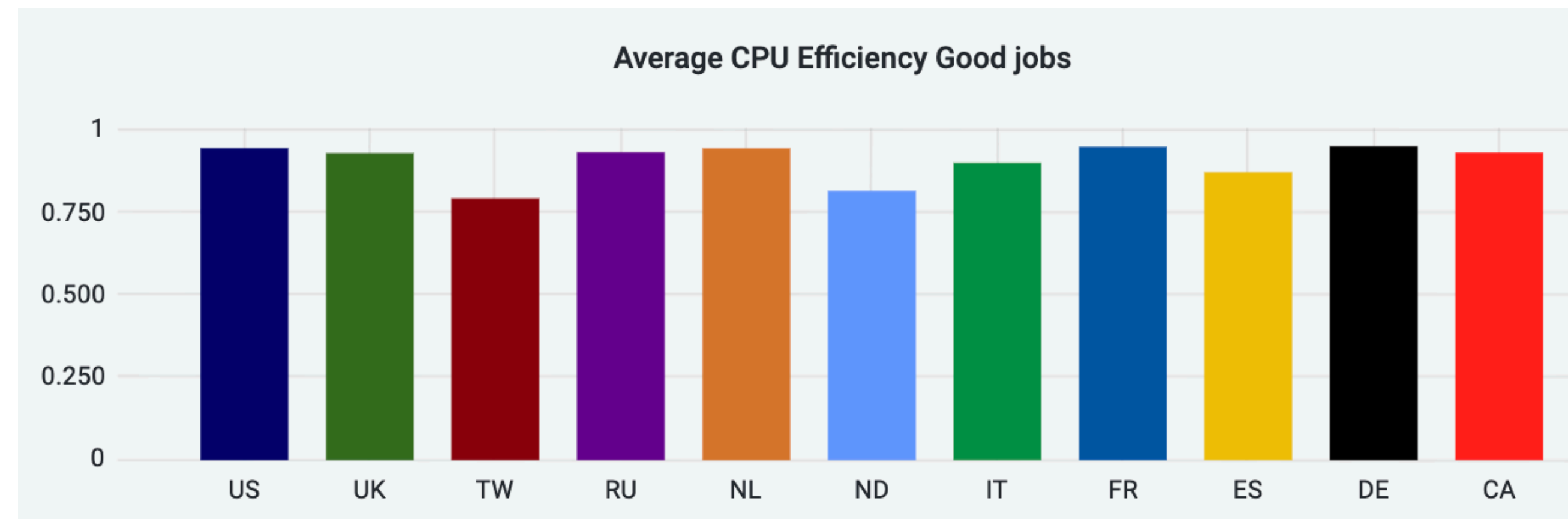
- 8% of walltime lost to failed jobs:
 - Dominated by problems with staging data to and from the WNs (or in direct access failures)

WallClock Consumption of Successful and Failed Jobs - Pie Chart



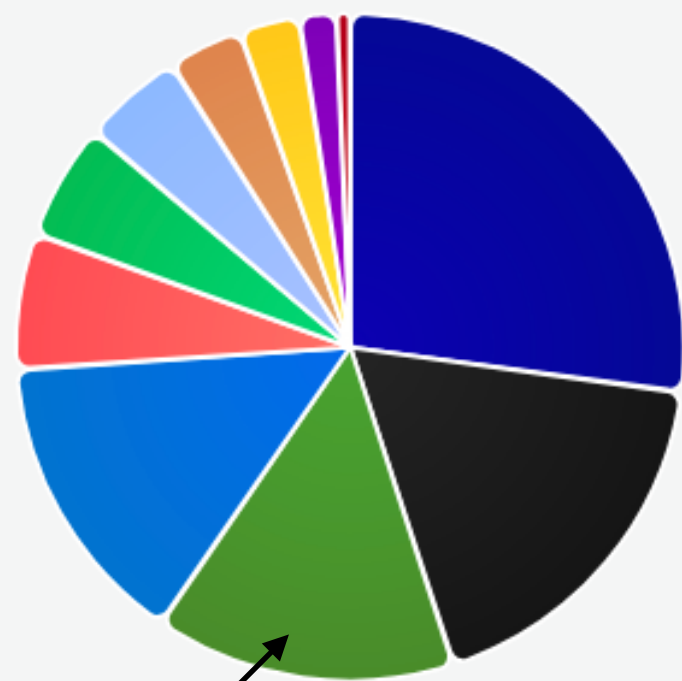
Cloud comparisons

- UK Contributes ~ 15% of total ATLAS compute across T1,2,3 grid sites.
 - Drops to 9% once HPC + cloud is included.



All (T1,2,3) Grid Sites

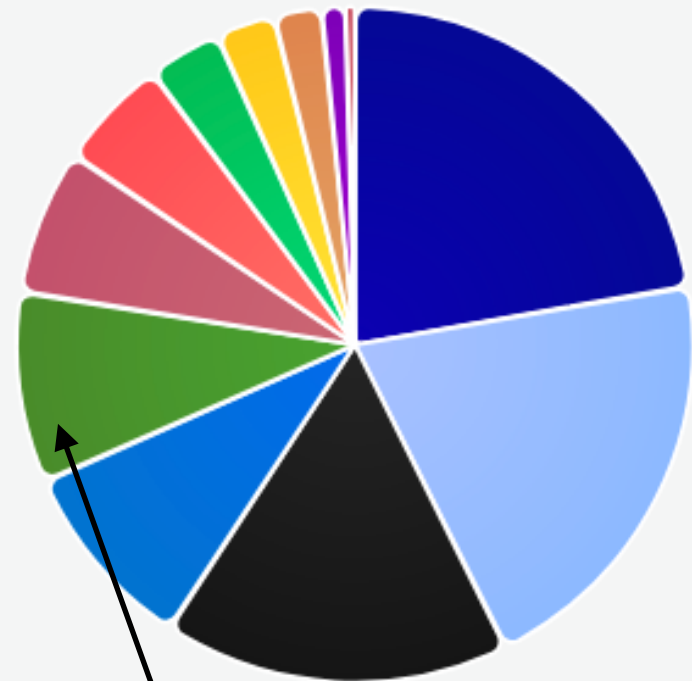
Slots of Running jobs (HS23)



UK ~ 15%

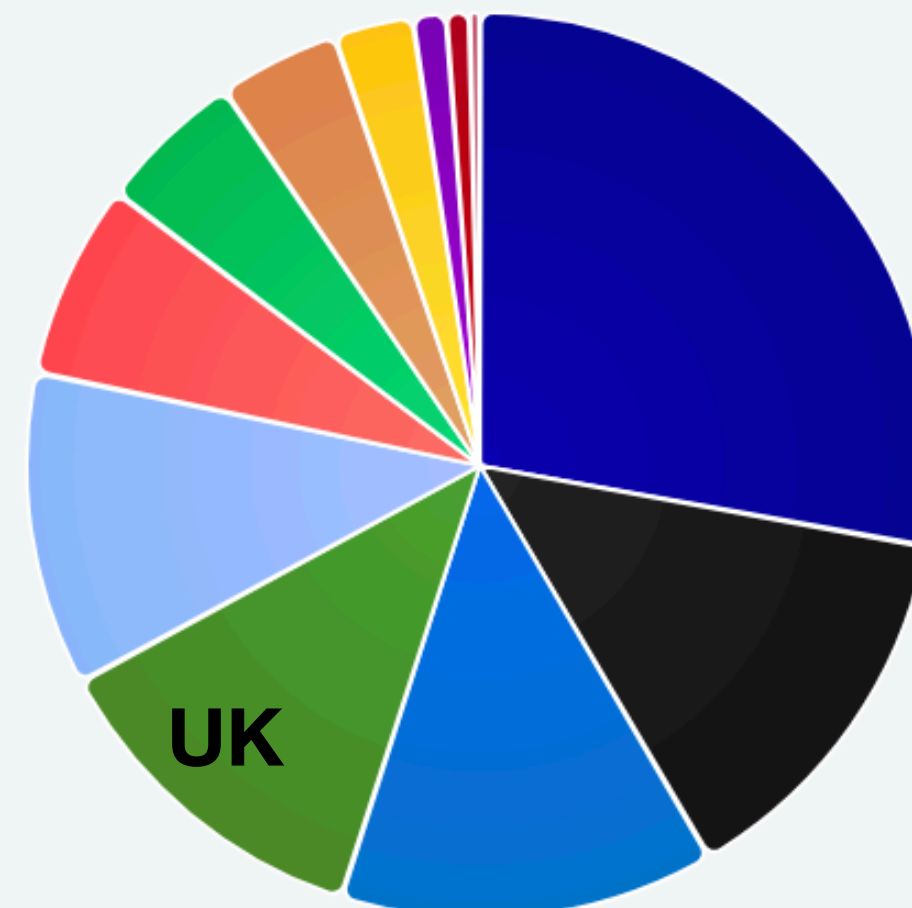
+ HPC, etc

Slots of Running jobs (HS23)

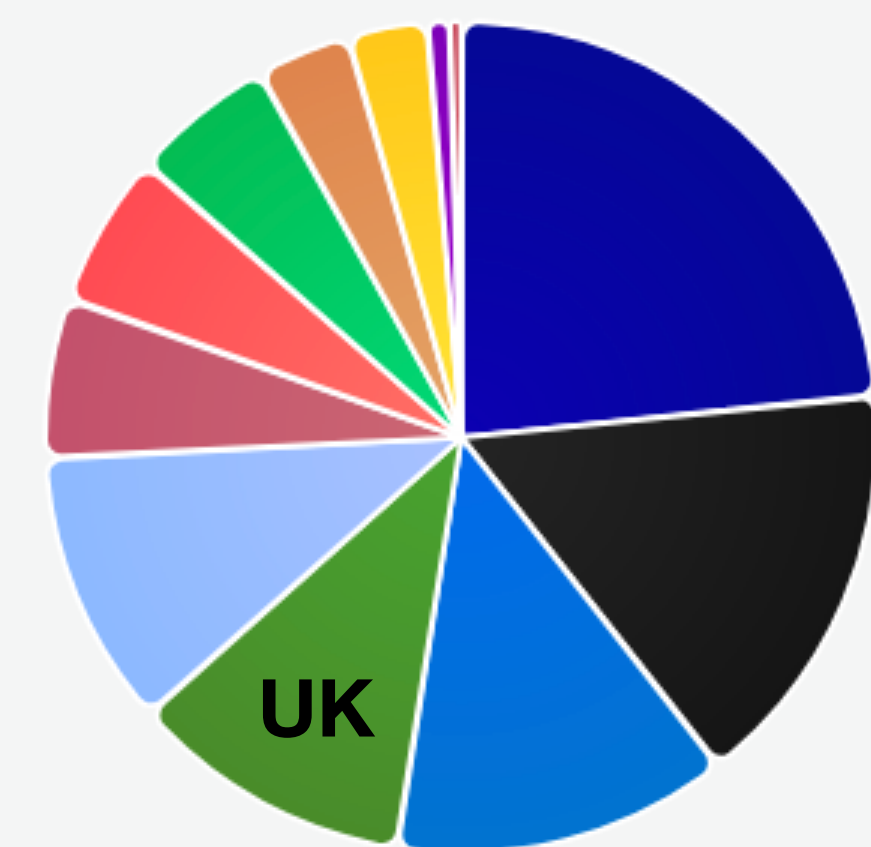


UK ~ 9%

NBytes Processed - in percentage



NBytes Produced - in percentage

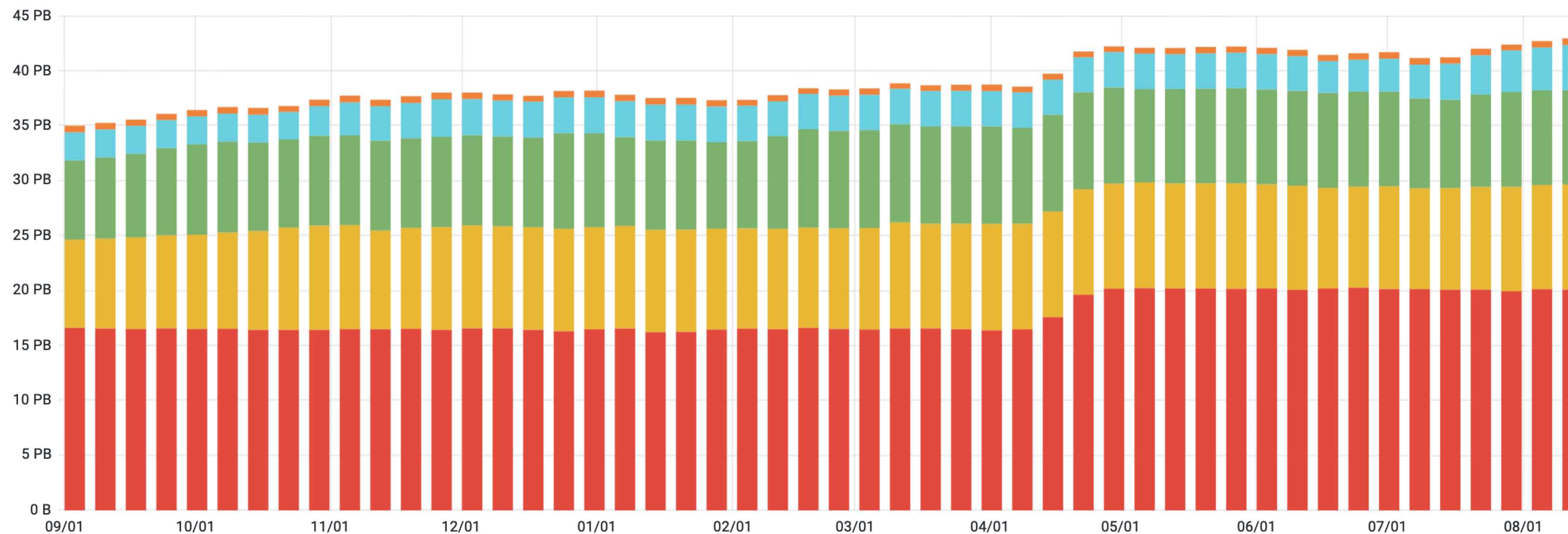


- UK processed ~ 300PB (*), producing ~ 24PB of output data in last year.

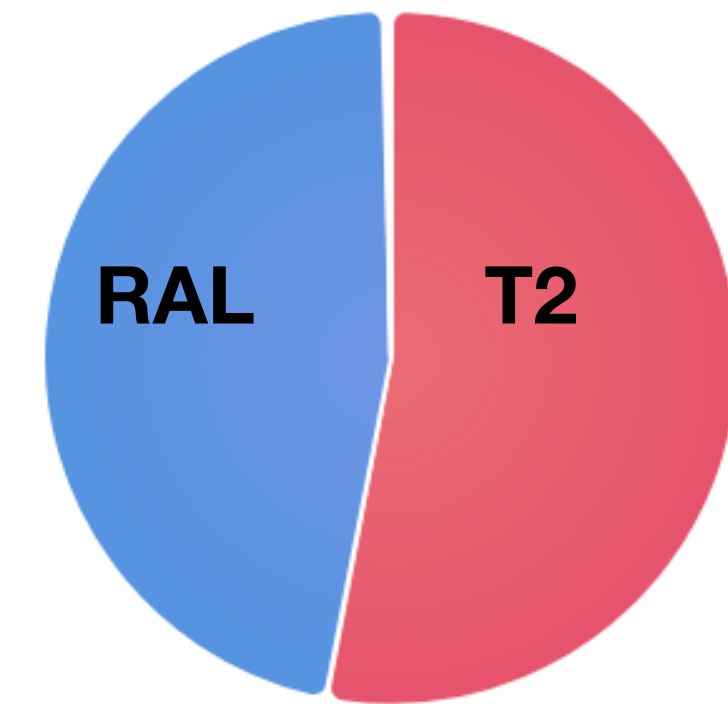
* – Over-estimated, as counts only total file sizes (and not e.g direct-io in user analysis)

Storage UK

- ~43 PB of (used) Datadisk in UK; split ~ evenly between T1 and T2 sites.
 - UK holds ~13% of disk data (by volume), 11% by number (~ 80M)
- Increased from ~34PB over last 12 months.



Volume by tier ▾



	max
UK-T1-RAL	20.3 PB
UK-NorthGrid	9.69 PB
UK-London-Tier2	8.94 PB
UK-ScotGrid	4.11 PB
UK-SouthGrid	616 TB

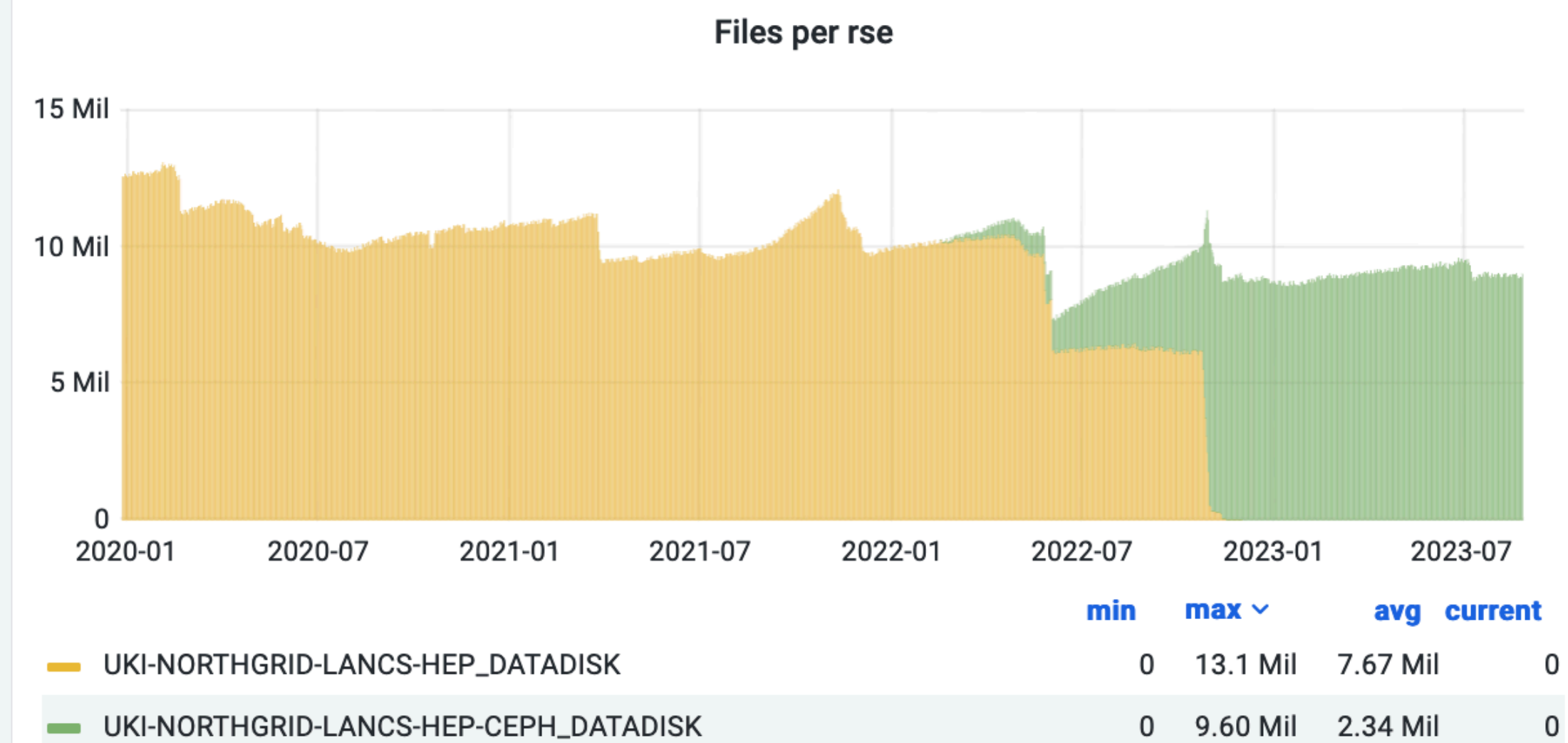
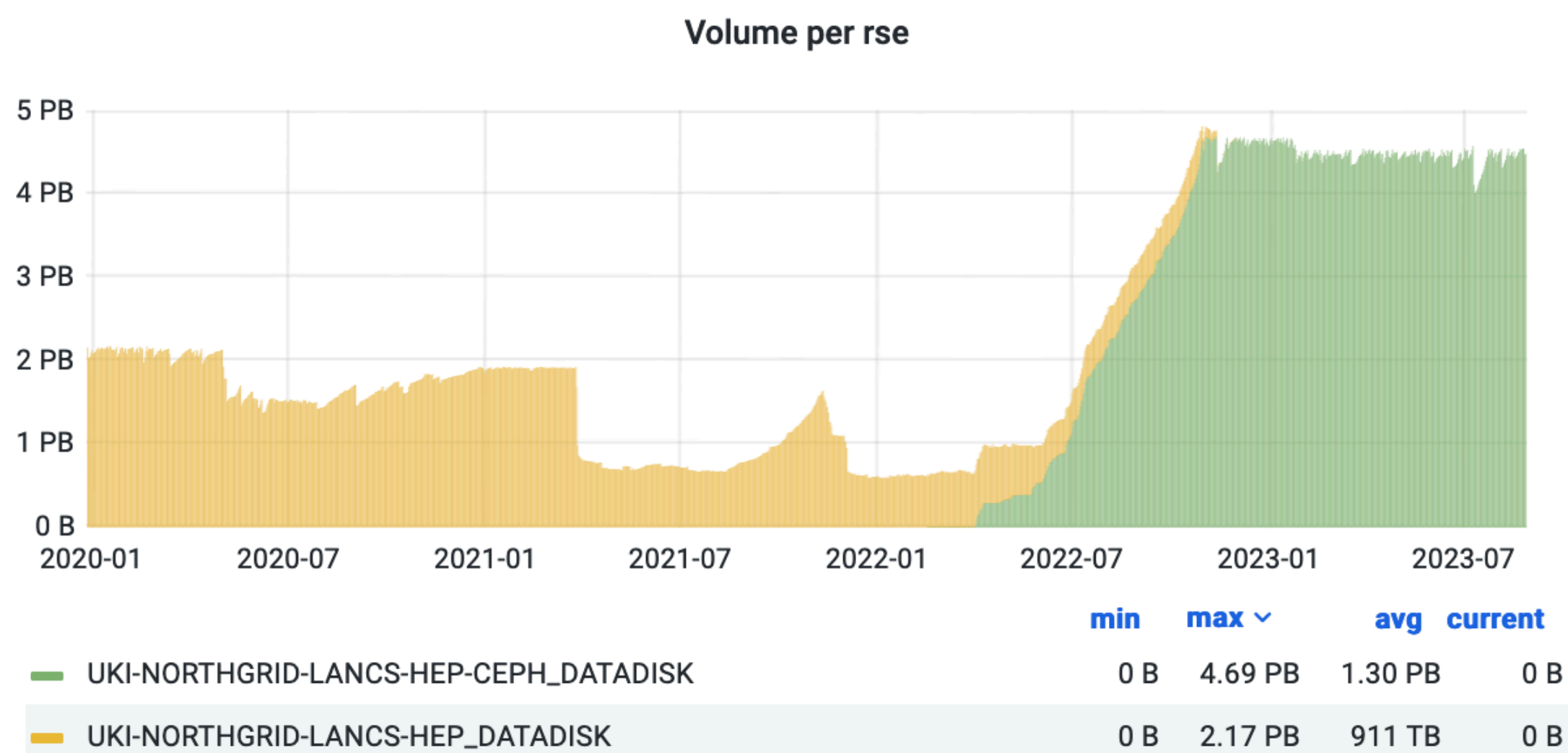
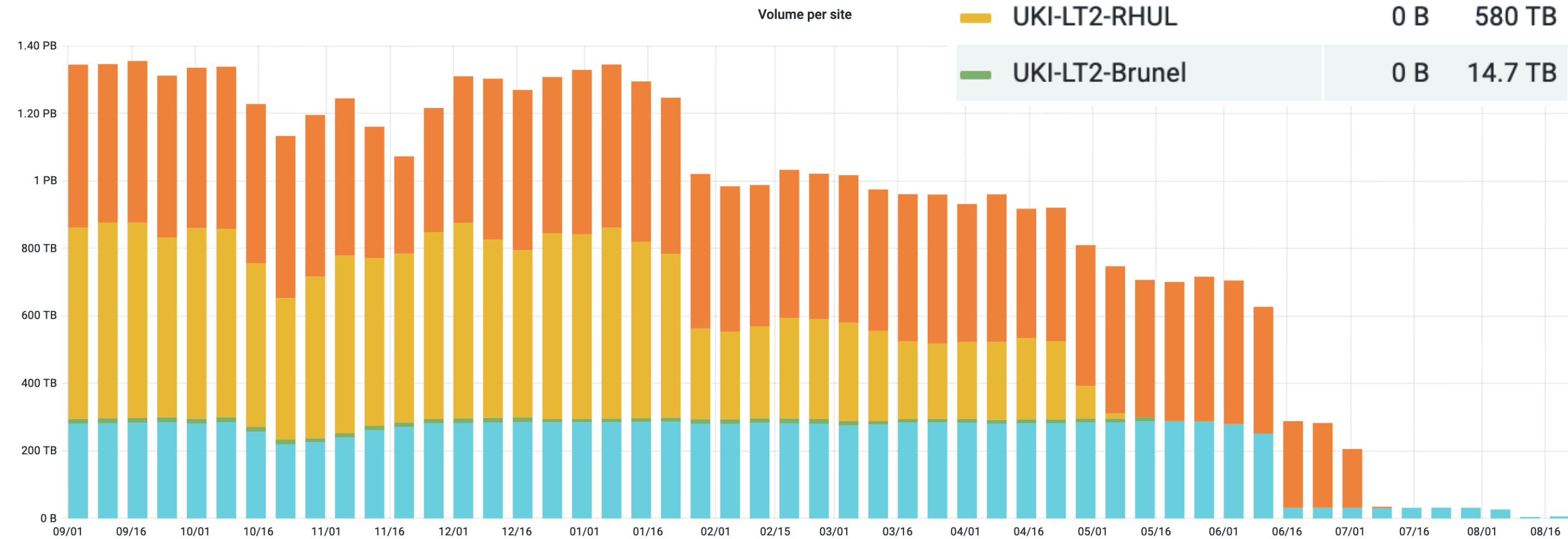
- RAL Tape:
 - 47PB of data currently written to tape
 - Continues to be used for 'day-to-day' operations with Data Carousel.



Storage decommissioning

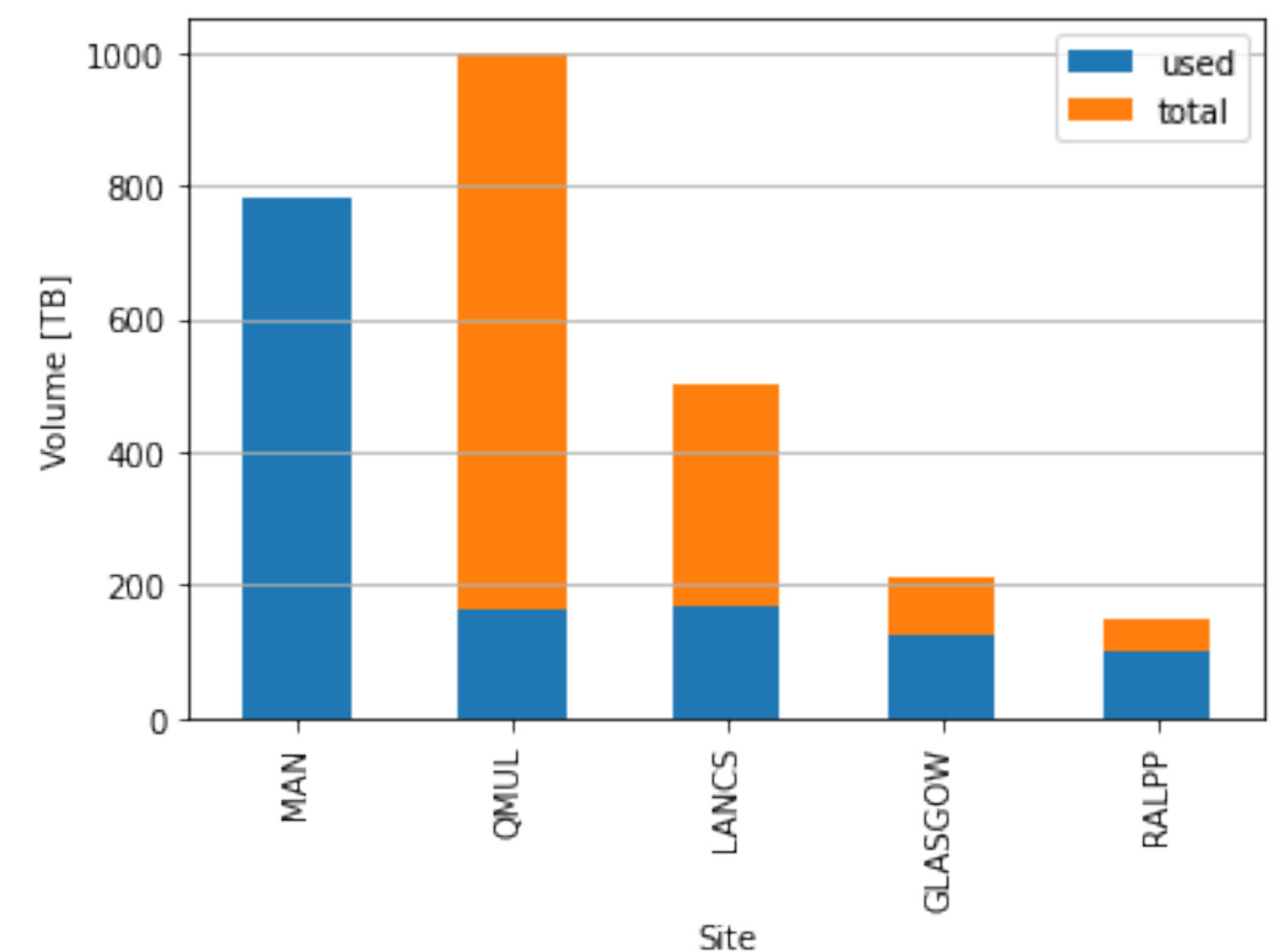
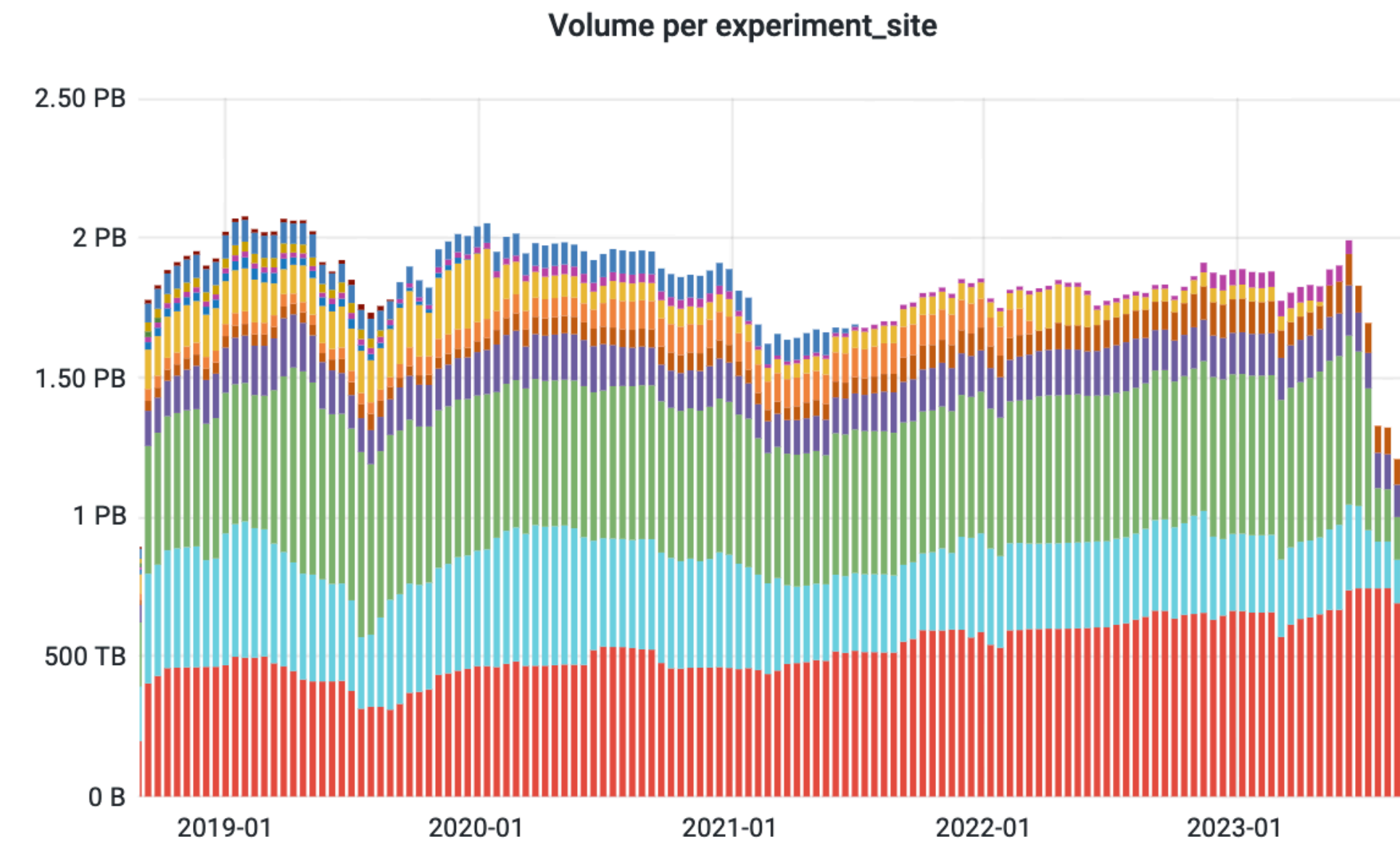
- Major operations in Storage decommissioning and migration in last year
- ~ 1.4 PB decommissioned from (Durham, ECDF, RHUL, Brunel)
 - (For Durham, part of storage migration)
- Manchester is final ATLAS UK site to migrate storage.
- Finalised migrations of other sites (e.g. LANCS)
 - (Good spring-clean opportunity to 'clear up' log files).

	min	max
UKI-SCOTGRID-DURHAM	4.41 TB	288 TB
UKI-SCOTGRID-ECDF	0 B	487 TB
UKI-LT2-RHUL	0 B	580 TB
UKI-LT2-Brunel	0 B	14.7 TB



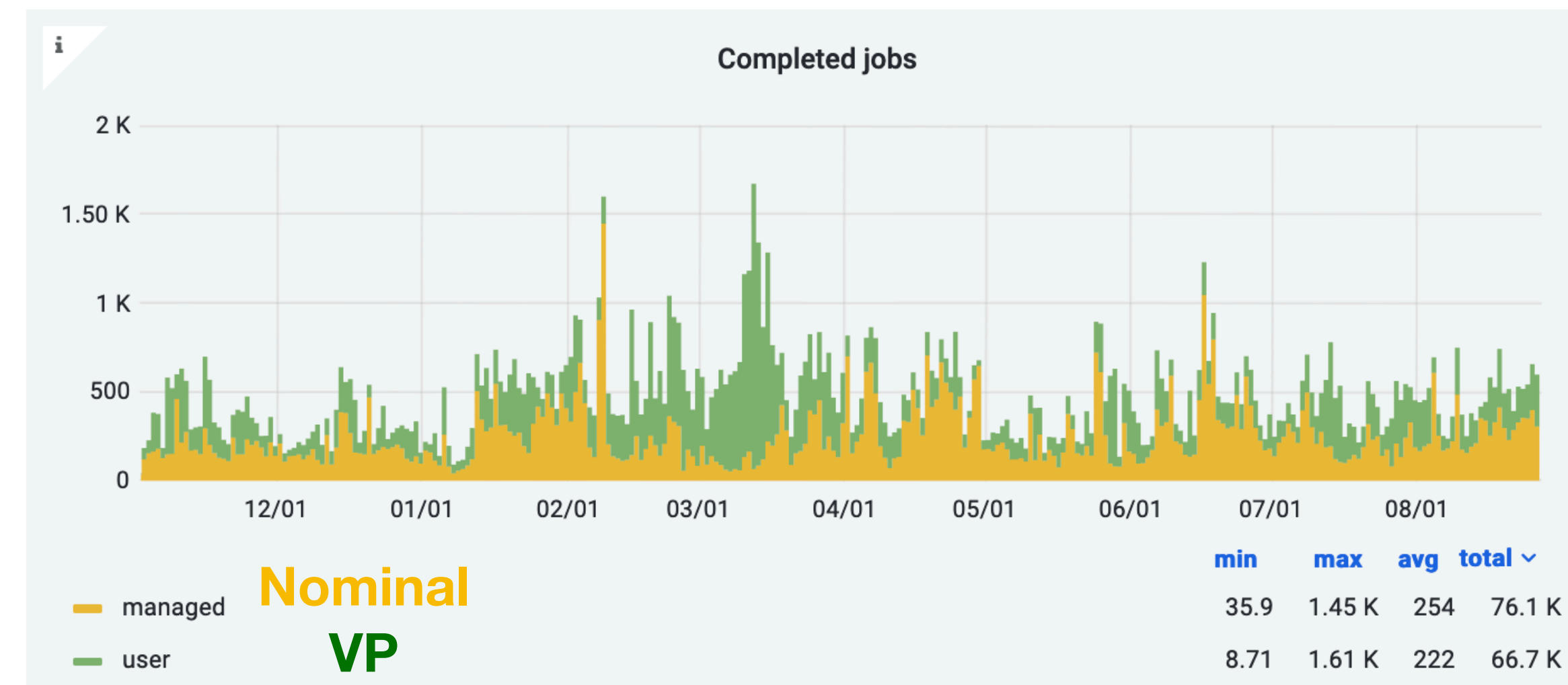
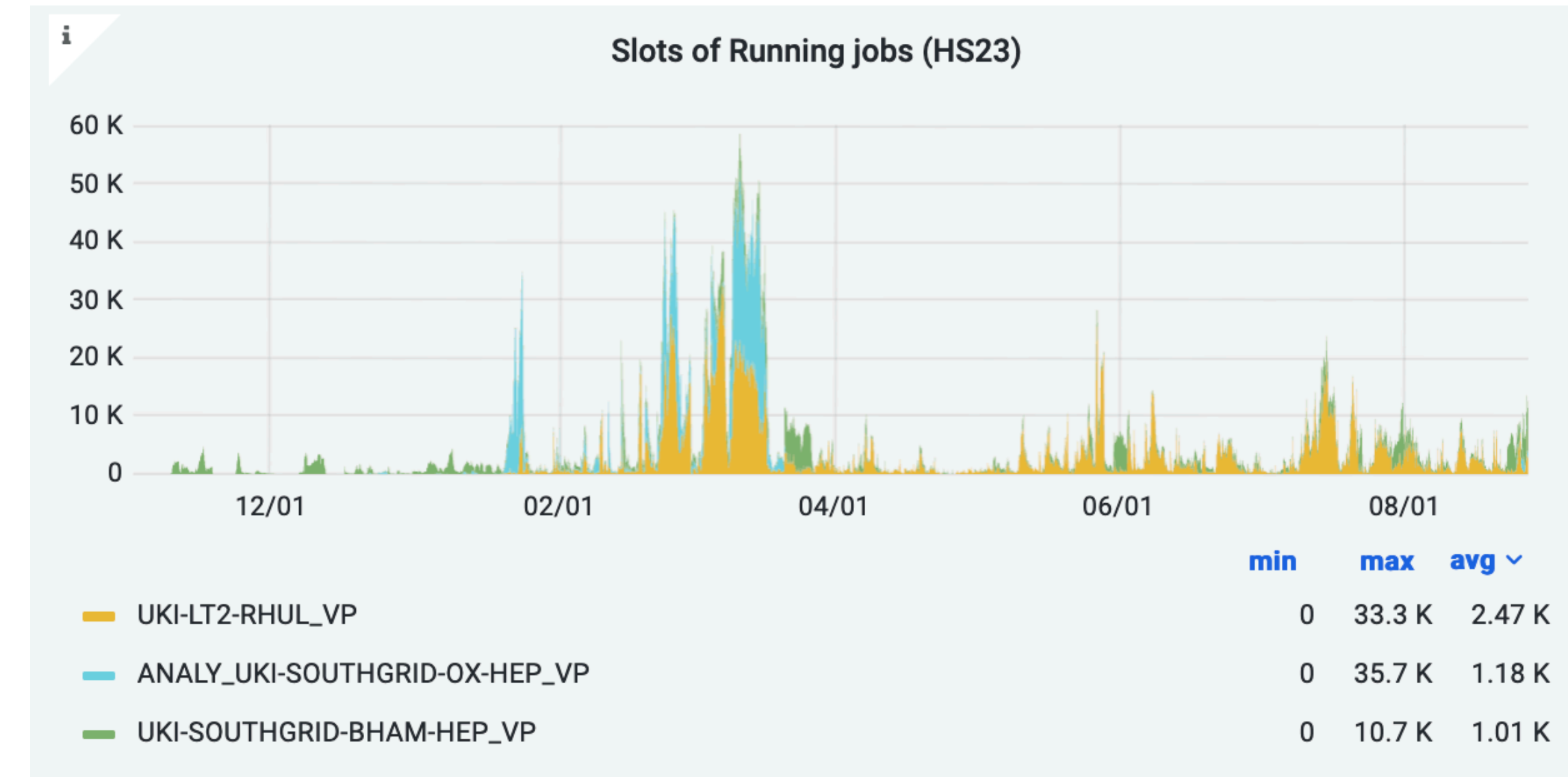
Localgroup disk

- Consolidation of Storage in UK also applies to LGD:
 - ~ 2.6PB made available to ATLAS UK users
- Currently available space doesn't (yet) appear critical overall:
 - Significant attempt to clean up older datasets recently
- Despite cleanup; Manchester remains most popular for users
 - QMUL - space available - awaiting DC upgrade,
- Need mechanism to 'nudge' users to other sites
 - Exploring quotas in rucio
 - More dynamic expressions in rucio rules (e.g. to move data to places which has space available)
 - Data locality for users will remain important however.
 - Balance needed between doing what's best for the user community, and fulfilling each user's desires.
- Communication with users:
 - Should ideally be based around atlas/uk group membership in IAM:
 - Group based 'admin' roles in IAM awaiting deployment (afaik).



XCaches

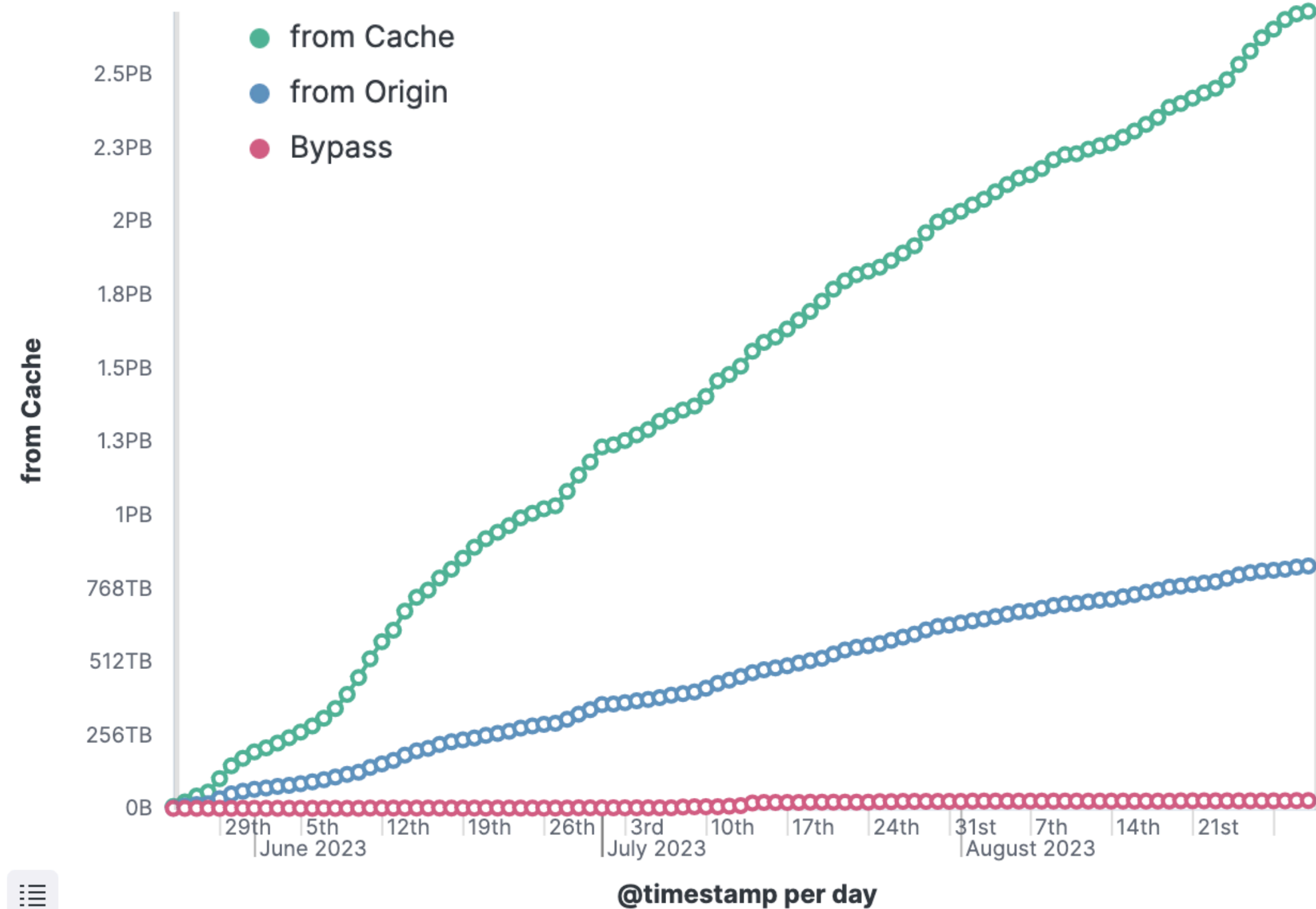
- Caching and data placement continues to cause interest
- OX, RHUL, BHAM, configured caches for Virtual Placement (VP)
 - Interest from Edinburgh
- VP – Use rucio to assign ‘virtual replicas’ at sites; aims to improve cache hit rates.
 - Only transfer the bytes that are needed for the workload
 - Configured for Analysis jobs
- RHUL very helpful in identifying issues (e.g. geo-location / blacklisted sites).
- ATLAS VP setup:
 - SLATE is recommended configuration;
 - Docker-compose available (with thanks to RHUL for improvements);
 - Manual configuration possible (OX); (complications due to updating from the previous configurations).
- Status:
 - Improving acceptance of VP within ATLAS as a de facto service;
 - Implies increased of small reads (and vector reads) from (any) SE.
- Still-to-do:
 - VP (analysis) and production queues currently both use the XCache:
 - Results in higher ‘churn rate’ of the XCache and lower hit efficiency
 - Need to test / move production queue to bypass the cache
- Still to demonstrate quantitatively the benefits of VP / XCache



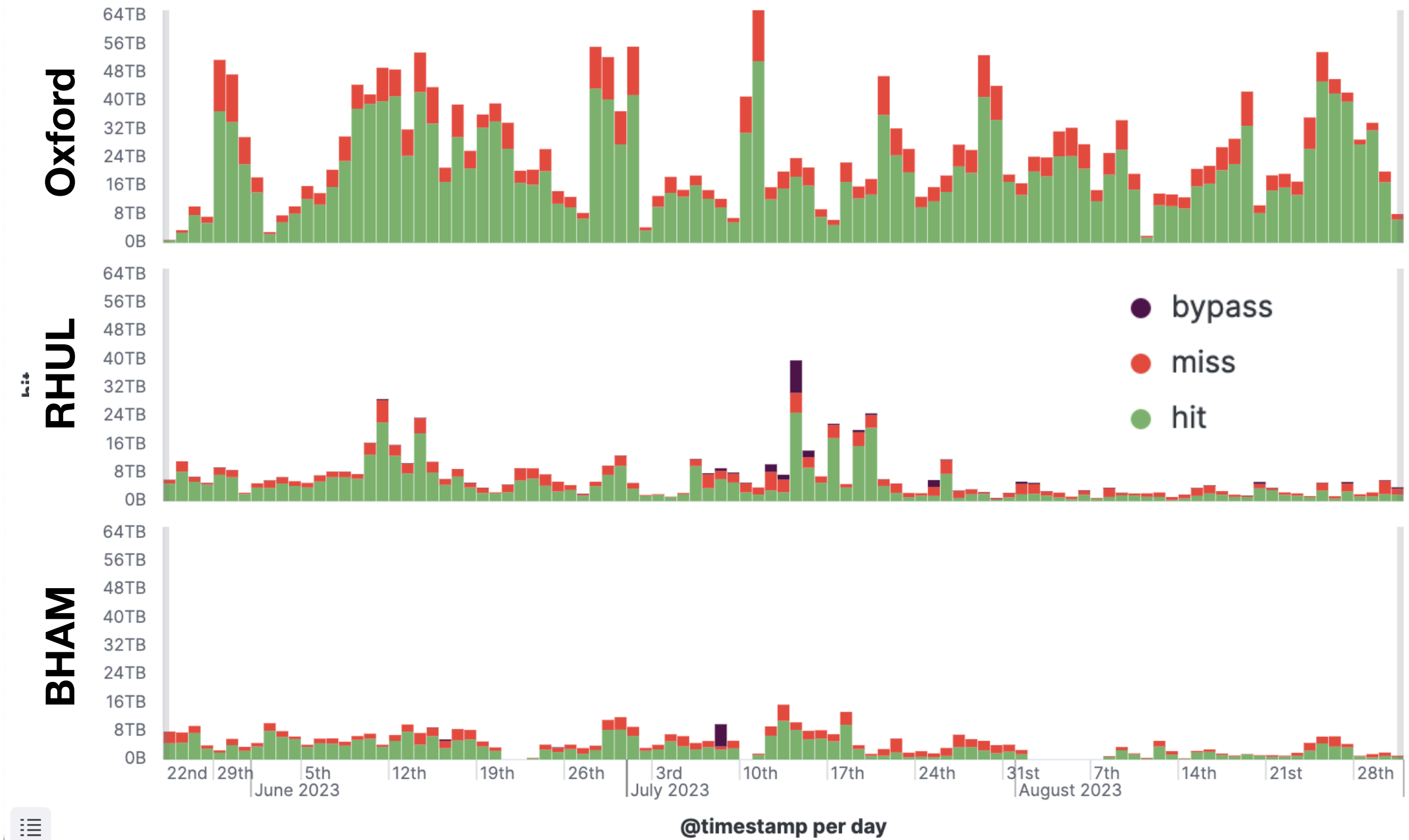
VP / Caching in use

- Xcache monitoring available at MWT2: [elasticsearch instance](#)
- Plots indicate over 2.5PB of data read from the Xcaches in UK, for ~800TB of actual external data transfer (TBC ...):

Cumulative data delivery

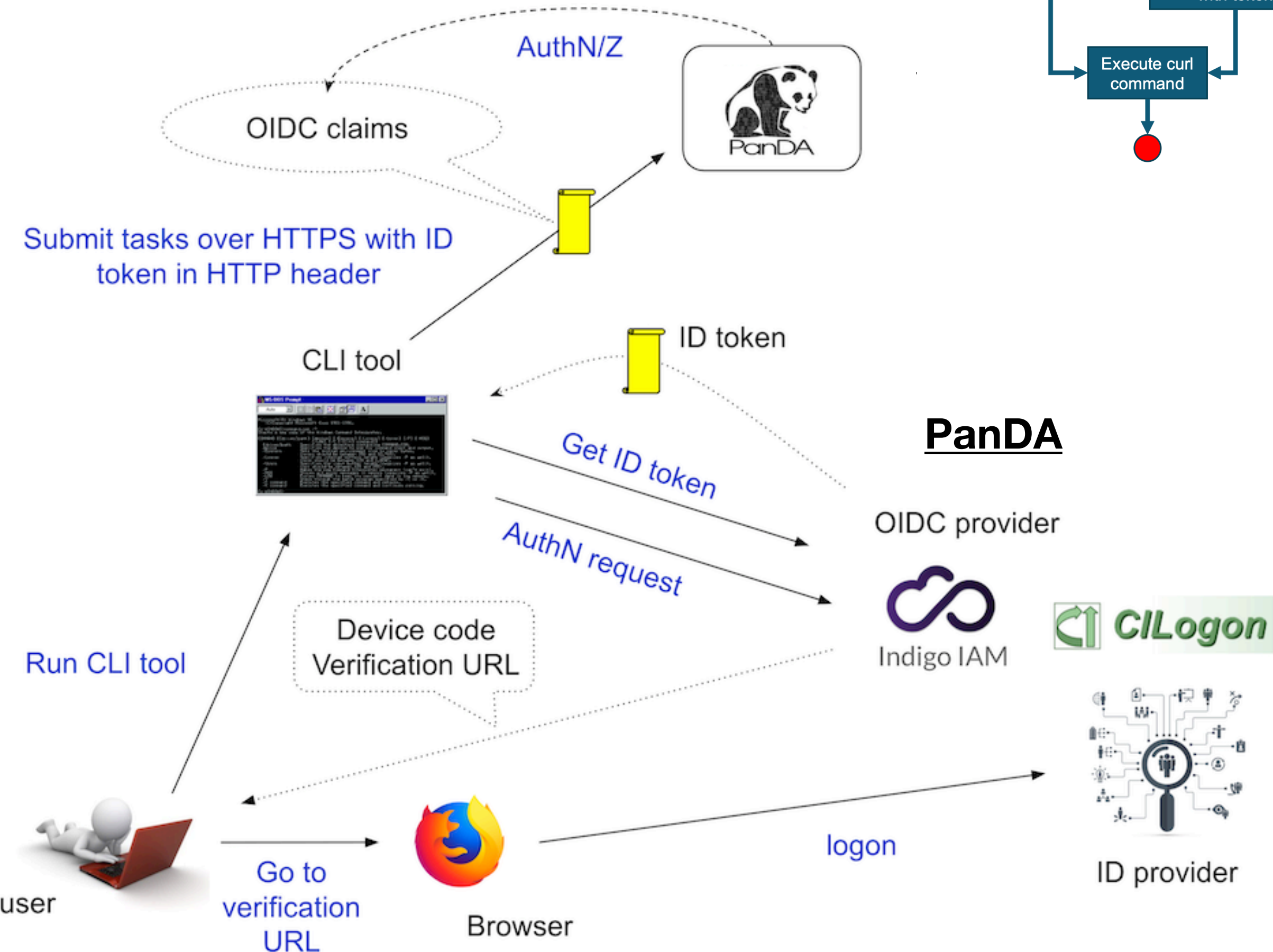


access bytes per site and hit/miss



Migration to Tokens

- ATLAS continues to move to token-based access (though not necessarily at the vanguard of initiatives)
 - x509 will remain for some time to come, together with tokens
- Storage:
 - Macaroon-based transfers already in production
 - Testing of IAM issued tokens proceeding
 - DPM doesn't support (sci/wlcg) tokens
 - Discussion remains on whether to use a 'common namespace' for path to files
 - e.g. /atlasdatadisk/rucio/xx/yy/file.root
 - To allow, e.g. a read token scoped to read from any storage
- Compute:
 - Support for HTCondor 9.0.x release series ended in May 2023
 - Deadline back in February 2023 for ATLAS sites to be ready for HTCondor 10 migration
 - Tokens required for HTCondor-CE (x509 no longer supported for submission)
 - New grid type "arc" supports only ARC-CE REST (still with x509)
 - All ATLAS jobs are submitted exclusively with tokens to HTCondor-CE
 - Possible to configure ARC-CE to accept job submission with tokens



Summary

- Observations:
 - UK holds significant expertise through it's personnel.
 - No obvious places where common tools are presented / disseminated (Confluence may be the solution?):
 - e.g. Lancs XRootD GitHub page is invaluable:
 - Links to monitoring / SRR reporting tools, ...
 - SLATE appears not be a preferred option for UK, but how to better leverage the expertise across sites with orchestration tooling?
 - Perfsonar - should be more useful, but only if all sites make concerted efforts to run / maintain it.
- DC24
 - Opportunity to see where things are currently 'creaking'
 - What do sites want to achieve from the challenge?
- Reminder:
 - Jyoti Prakash Biswal continues as ATLAS T1 Liaison (@50% level)
 - With support from UK Cloud Squad and sites
 - Post for for full FTE liaison now advertised (8 Sept deadline)
 - (I will still be around at low-level for some time)
 - <https://mattermost.web.cern.ch/gridpp/channels/atlas-support-cloud-uk>
- Thank you for helping support me in this role; educating me, and ensuring the successful operation of ATLAS distributed computing

<https://tinyurl.com/atlasjobt1>

