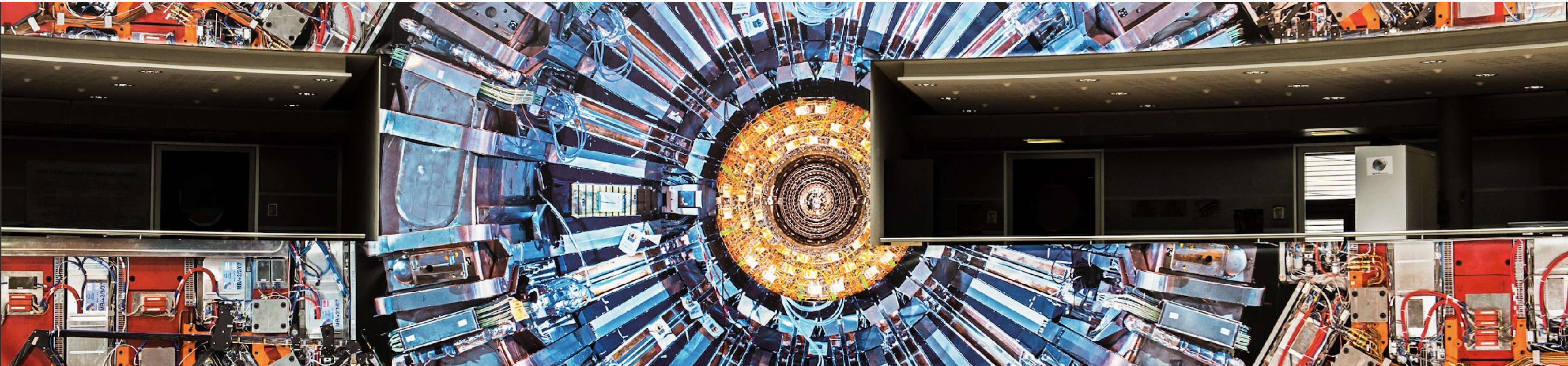
CMS Update

J. Letts (UCSD), D. Piparo (CERN) - LHCOPN/LHCONE Meeting #49 - October 25, 2022



CMS Networking Strategy

CMS thinks strategically about Networking:

- The “Third Pillar” of computing resources along with processing and storage
- The interplay between them is something that CMS exploits with a flexible computing model. Some examples:
  - Since Run 2, pileup pre-mixing libraries are streamed remotely with XRootD:
    - Avoids the need to have local copies of the PU libraries, conserving disk space
  - “StepChain” model of workflow management: conceived for the HPC use case, now applied almost universally for production since 2020:
    - Steps in Monte Carlo event production: Generation, Simulation (Geant4), Digitization, Reconstruction, and creation of “slim” data products (MiniAOD & NanoAOD)
    - Workflows handle all steps of the chain, and intermediate data products are not copied over the network but serve as the input for the next step in the chain and run locally.

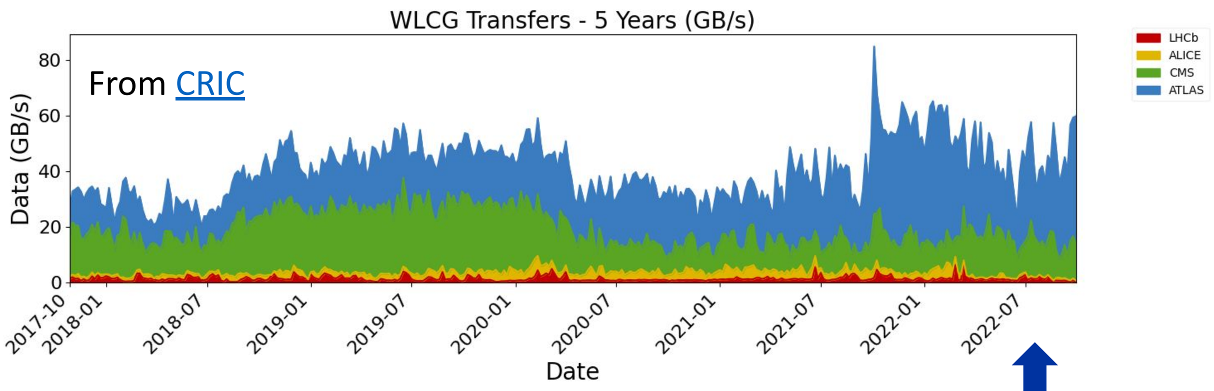
The flexibility to “trade” one resource type for another allows us to adjust to changes in cost or availability.

Current CMS Networking Use Cases

- During data taking, transferring a second, working copy of the RAW data to our six Tier-1 sites (Rucio & FTS)
  - Driven by the RAW event size: We are working to improve our understanding the Phase-2 RAW size.
  - Especially heavy during heavy ions running
  - CMS adopted Rucio as the data transfer software in late 2019:
    - Opportunity for us to revisit our data placement policies to reduce network usage.
- Monte Carlo production:
  - Pre-mixed pileup library streaming (XRootD) - unscheduled
- Bulk data transfers for
  - Production (input and output samples, with Rucio & FTS) and
  - Analysis needs (Rucio & FTS, as well as XRootD)

How will this evolve in the future (Run 4 and beyond)?

- Other heavy users of network (e.g. SKA) will enter the arena. Lots of new technologies to manage transfers in a more competitive and constrained environment.
- Inclusion of storageless sites, federated storage, caching: the landscape is becoming more and more heterogeneous.
- A hierarchy of network bandwidth capabilities is the reality: We are a global Collaboration and different regions moved away from the original tiered MONARC model at different rates.
  - A hybrid system will likely need to be sustained throughout the next decade, regardless of advances exploiting new technologies and capabilities in certain regions.



CMS Offline Software and Computing enables a *global* physics program.

We are closely monitoring our network usage during Run 3. We'll learn a lot from 2023.

HL-LHC R&D

There are many interesting directions of technological development which will affect how CMS uses ***and manages*** the network (as well as CPU and storage) over the next decade.

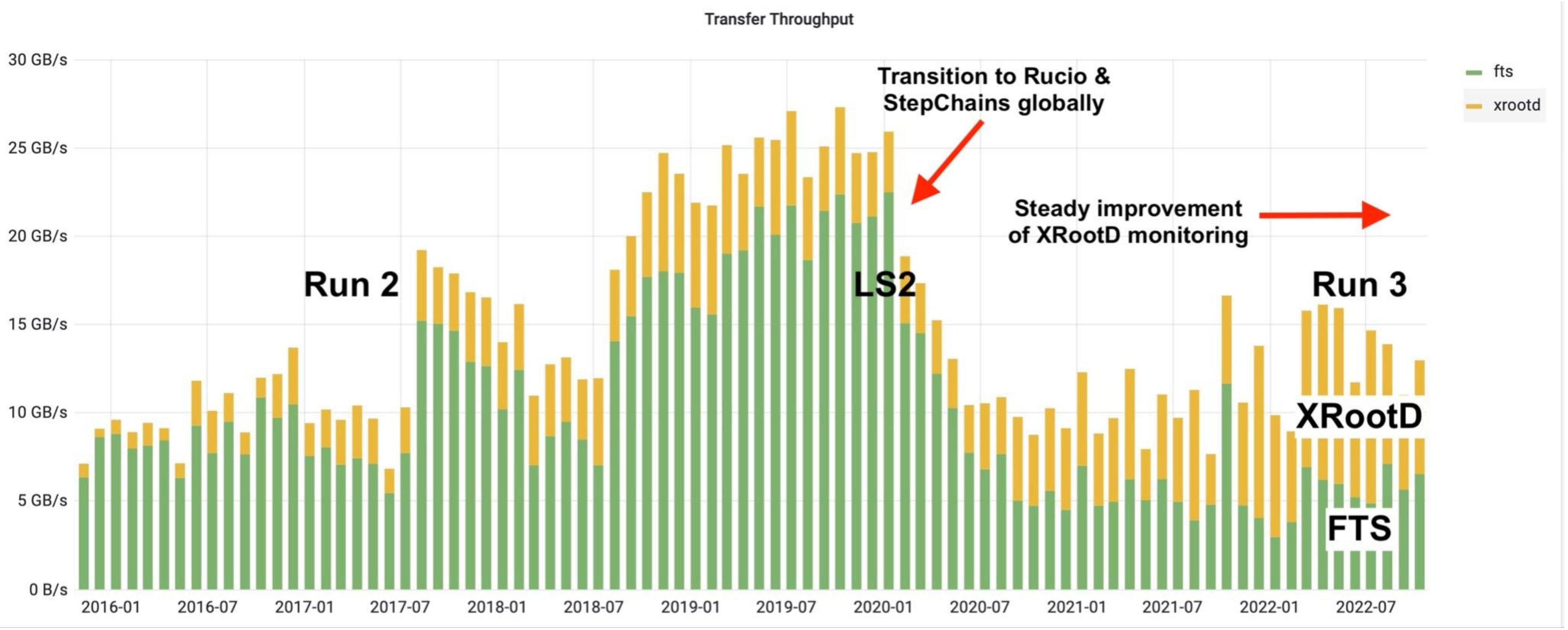
CMS members are involved in R&D and technology evolution, some examples:

- Token-based authentication transition
- Federated storage, caching
- Analysis facilities and software tools
- SDNs: NOTED, Rucio and SENSE integration:
    - During the next round of data challenges, CMS is planning to use the opportunity to run the SENSE & Rucio prototype at a couple of sites on a production scale.

CMS has started the process to write a Conceptual Design Report for software and computing for HL-LHC: How we manage data transfers and the network will be an important part of that planning document, as well as estimating the networking needs of the HL-LHC CMS computing model.

Seven Years of CMS Data Transfers

- The transition to Rucio gave us an opportunity to review our data transfer policies.
- Improvement of the monitoring (especially for unscheduled XRootD transfers) is key.
- Next year's usage (full Run 3 production year) will be very interesting: 1/3 of the way to Run 4 scales!



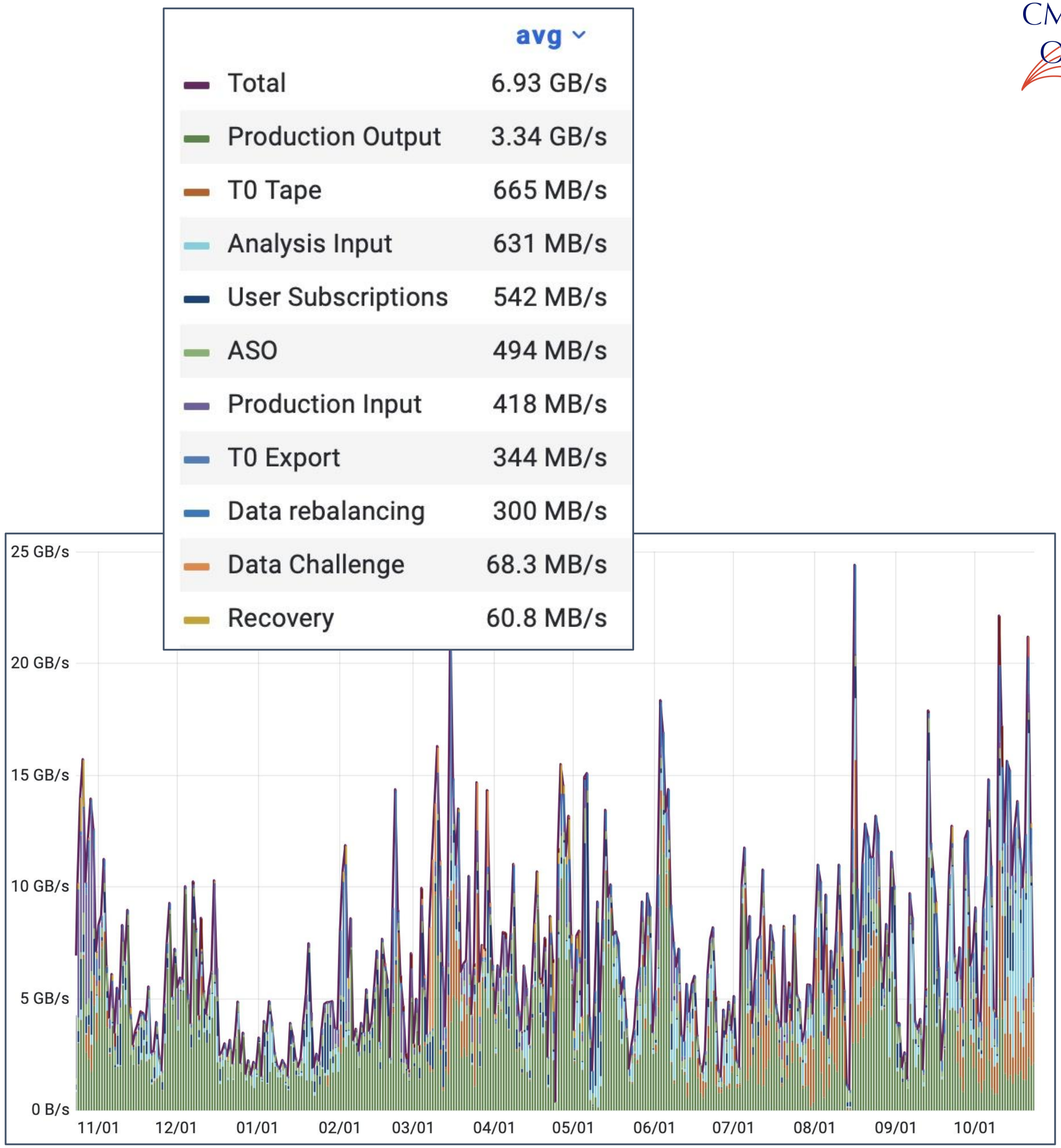
Source: [WLCG Transfers]

Monitoring Improvements

Better monitoring is key to progress:

- We are encouraged by recent improvements to the XRootD monitoring at U.S. sites and look forward to the implementation in other regions.
- Packet marking is another thing we are very interested in, but lack effort to really drive it.
  - CMS *is* dedicating more effort in this area.
  - If we can understand our scheduled and unscheduled transfers with more granularity, we can likely improve our usage of networking.
  - Base categorization already exists in CMS in-house FTS monitoring:

“If you can monitor it, you can fix it”



Assessing Future Needs

As a baseline, WLCG & experiments did back-of-the-envelope estimates of HL-LHC needs by extrapolating Run 2 network usage by the experiments to PU=200 scales. A lot has changed since then:

- Run 4 start has slipped from 2027 to 2029, with the first full production year 2030 with PU=140 instead of PU=200.
- PU=200 will be reached in Run 5, more than a decade from now.

Still, it's a very good starting point:

T1	LHC Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Flexible Scenario in 2027	Data Challenge target 2027 (Gbps)	Data Challenge target 2025 (Gbps)	Data Challenge target 2023 (Gbps)	Data Challenge target 2021 (Gbps)
CA-TRIUMF	200	400	100	60	30	10
DE-KIT	600	1200	300	180	90	30
ES-PIC	200	400	100	60	30	10
FR-CCIN2P3	570	1140	290	170	90	30
IT-INFN-CNAF	690	1380	350	210	100	30
KR-KISTI-GSDC	50	100	30	20	10	0
NDGF	140	280	70	40	20	10
NL-T1	180	360	90	50	30	10
NRC-KI-T1	120	240	60	40	20	10
UK-T1-RAL	610	1220	310	180	90	30
RU-JINR-T1	200	400	100	60	30	10
US-T1-BNL	450	900	230	140	70	20
US-FNAL-CMS	800	1600	400	240	120	40
(atlantic link)	1250	2500	630	380	190	60
Sum	4810	9620	2430	1450	730	240

Table 2: data challenge target rates.

The CDR process over the next year should clarify the CMS needs more precisely.

CMS in the Network and Data Transfer Challenges

CMS participated in the March 2022 data challenges: Transfer paths of interest met (or were close to, in some cases) the target rates transfers to *tape*, less than network bandwidth by construction.

End-to-end transfer rates are currently not limited by network wire speed. Lots of other factors to consider, from bandwidth to and from storage, software, disk space …

- CERN-to-T1 tape rates <1/2 total network throughput targets. (e.g. 16Gbps:40Gbps)
- This is one important aspect of the transfer capability that the experiment really cares about.

CMS site	March 2022 rates (GB/s)	Target Rates (GB/s) *(to TAPE)*
CTA	5.72	3.2
KIT	0.59	0.29
PIC	0.58	0.15
Fermilab	1.97	0.73
RAL	0.8	0.29
IN2P3	0.59	0.29
JINR	0.50	0.55
CNAF	0.32	0.37

Conclusions

- Networking is a strategic concern in an experiment like CMS with a flexible computing model.
- Improving monitoring is key to improving usage!
  - Making the XRootD monitoring more complete (e.g. of unscheduled traffic)
  - Packet marking for in-depth understanding of usage of both scheduled and unscheduled traffic.
- We encourage and participate in R&D and other improvements:
  - Token-based authentication transition
  - Federated storage, caching
  - Analysis facilities and software tools
  - SDNs: NOTED, Rucio and SENSE integration
- WLCG Network Challenge goals are a good baseline for HL-LHC needs
  - Take into account the new LHC schedule & PU profiles for a now more realistic timeline
  - CMS will investigate its long-term needs more precisely as part of the CDR process over the next year.

Finally we would like to thank our partners for their collaboration: OSG, ESnet (especially for the upgrades of the transatlantic link), other NREN providers, LHCONE / LHCOPN, CERN, WLCG, ATLAS & other experiments, the Rucio development team, EGI, national laboratories … and many others.