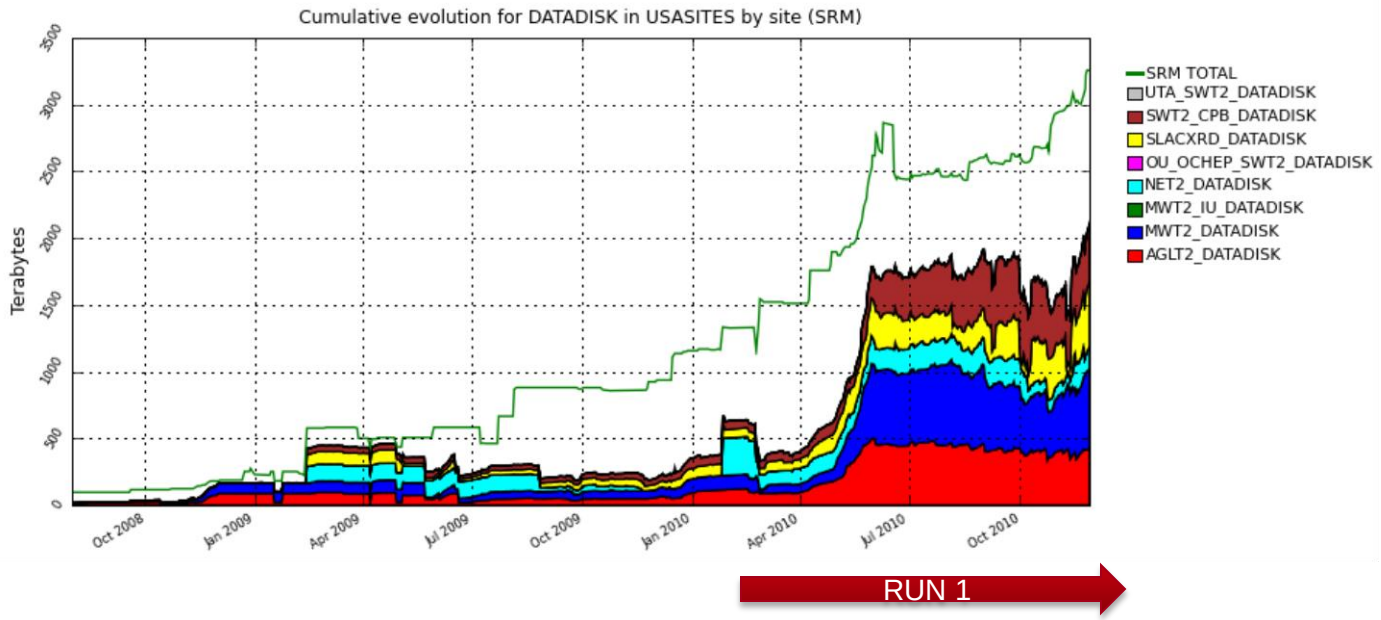
Concerns from Experiments ATLAS

Richard P Mount
SLAC National Accelerator Laboratory

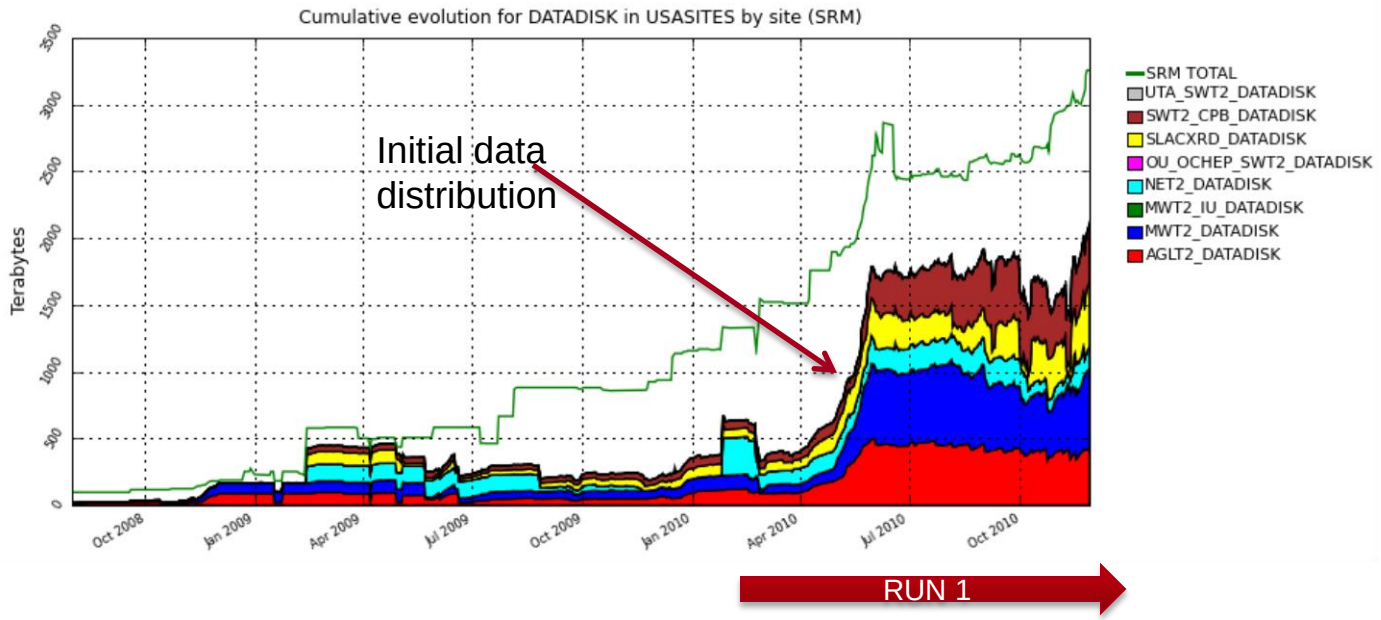
ATLAS Computing at the Start of Run 1

- Distribute data to WLCG sites according to a policy approved by a committee:
  - 2.5 disk copies of the ESD at Tier 1s
  - 2.5 disk copies of the ESD at Tier 2s
  - 2 primary + 8 secondary disk copies of AOD
  - etc.
- Send Grid jobs to the data

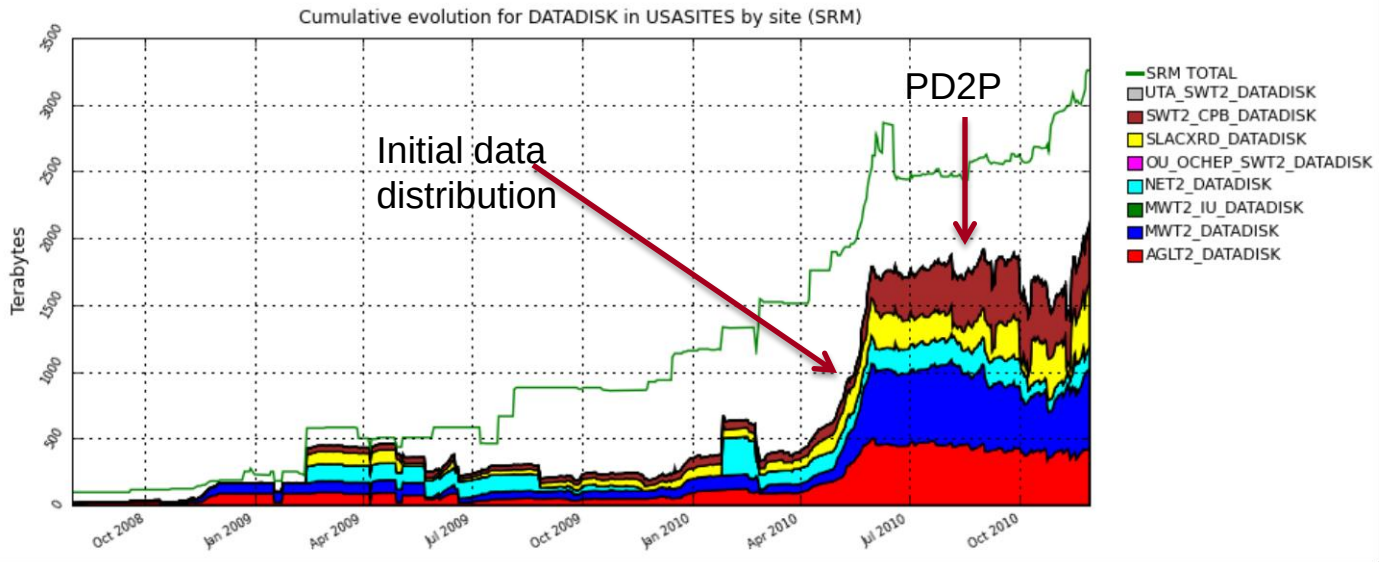
ATLAS Disk Space Usage – Early Run 1



ATLAS Disk Space Usage – The Crisis



ATLAS Disk Space Usage – The Solution

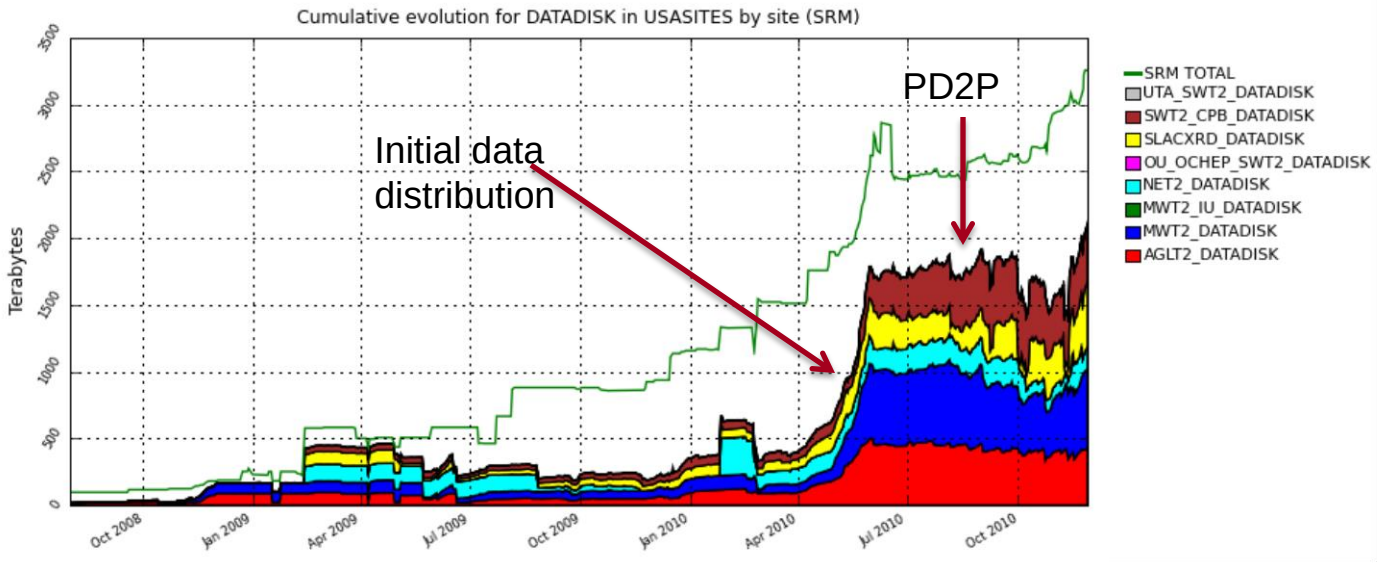


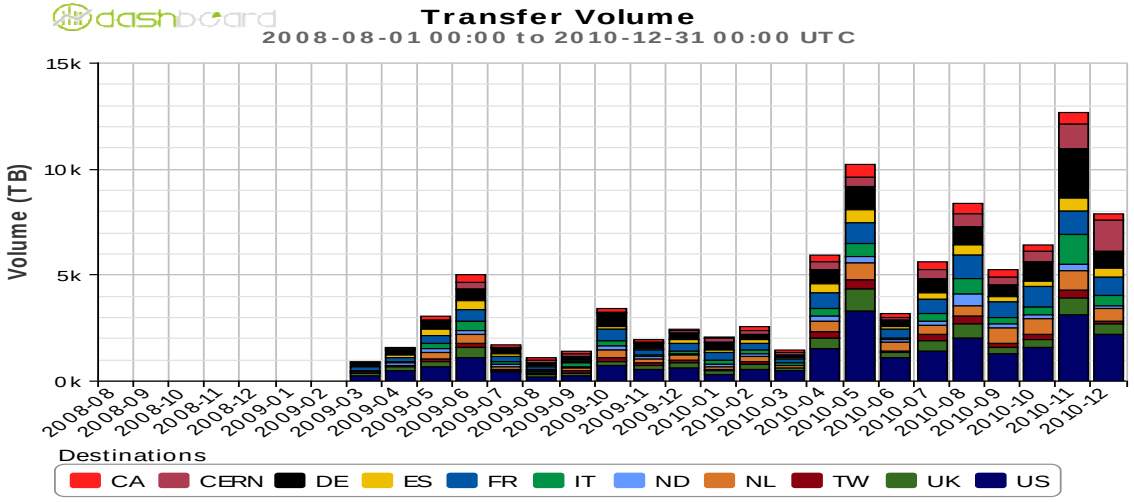
PanDA Dynamic Data Distribution (PD2P):

- Suppress most policy-based replication to T2s
- Replicate datasets to T2 when they are in demand at T1s
- Re-broker jobs from T1 queues to T2 queues when the data they need arrives at T2s

- Physicist experience: nobody noticed!

ATLAS Disk Space Usage – Network Impact





PD2P: Dataset Reuse in 2012 – Qualified Success

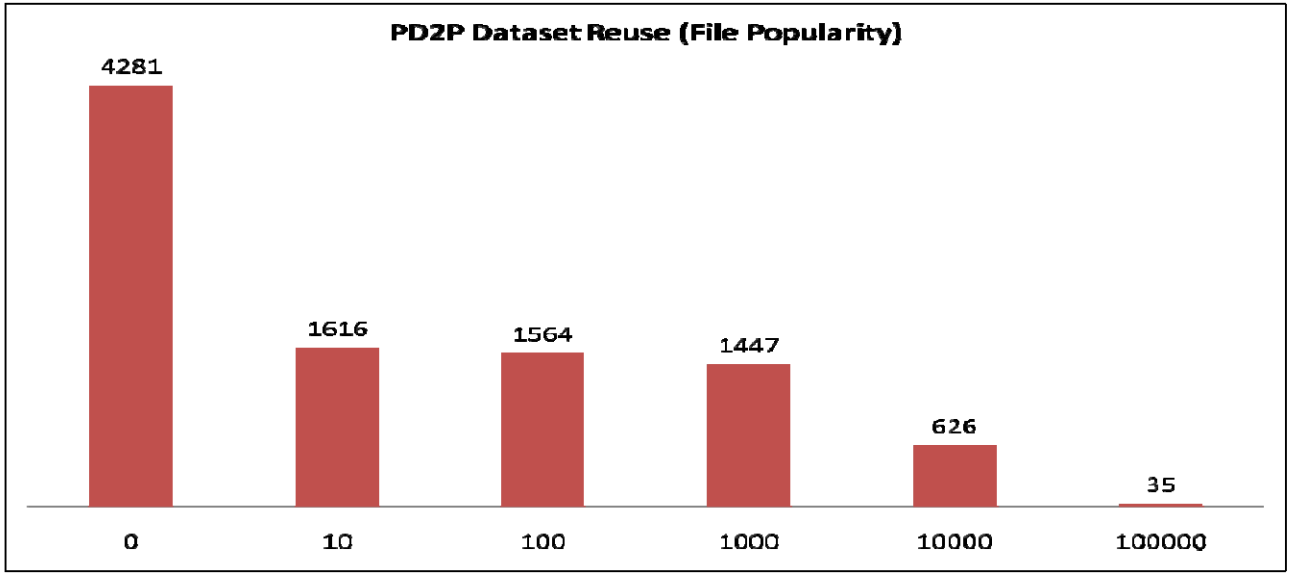


Figure 4. Frequency of Reuse for PD2P datasets

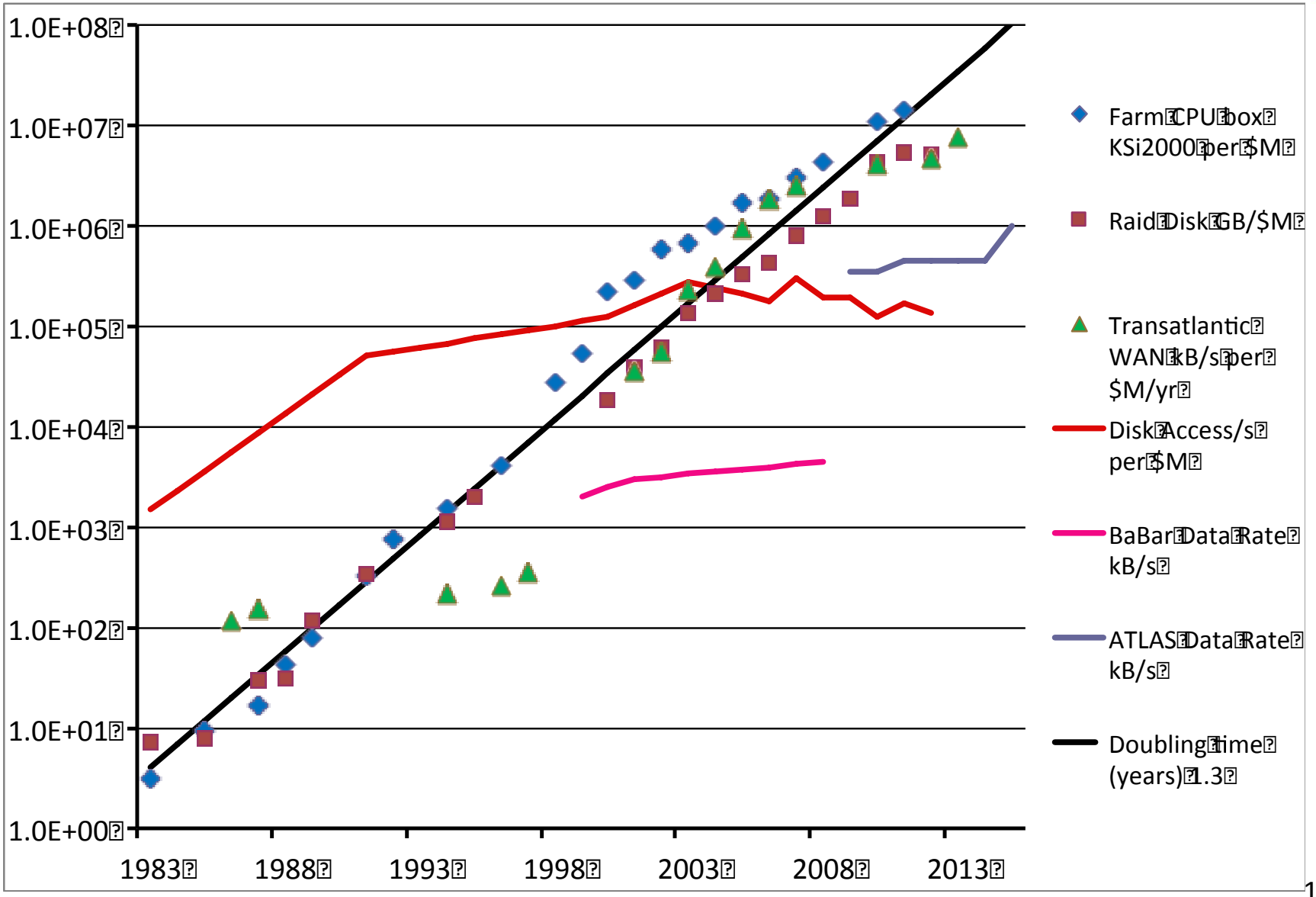
Recent ATLAS DDM Operations

- Life without ESD
- Regular T1 disk crises
- T1 disks almost filled with data marked as “primary”
- October 2013 C-RSG
  - “whilst we welcome the more aggressive policy for the deletion of unused data, we think that, given the volume of unread data and the cost of disk, unused space could be recovered more aggressively and more of the disk-resident data at T2s placed under ‘load on demand’ management.”
- To survive (not just pacify the C-RSG) we have to aggressively delete (or not replicate) little-used data

Disk Cost – the Driver of Network Use

- For 30 years we saw bytes/CHF rise by a factor of two every 18 months or less.

The Past: Exponential growth of CPU, Storage, Networks



Disk Cost – the Driver of Network Use

- For 30 years we saw bytes/CHF rise by a factor of two every 18 months or less.

- Disk seems to have reached the end of the road with current technology. New technologies are said to be almost production ready but:
- Expect modest growth (factor 2 every 4 years?) in bytes/CHF

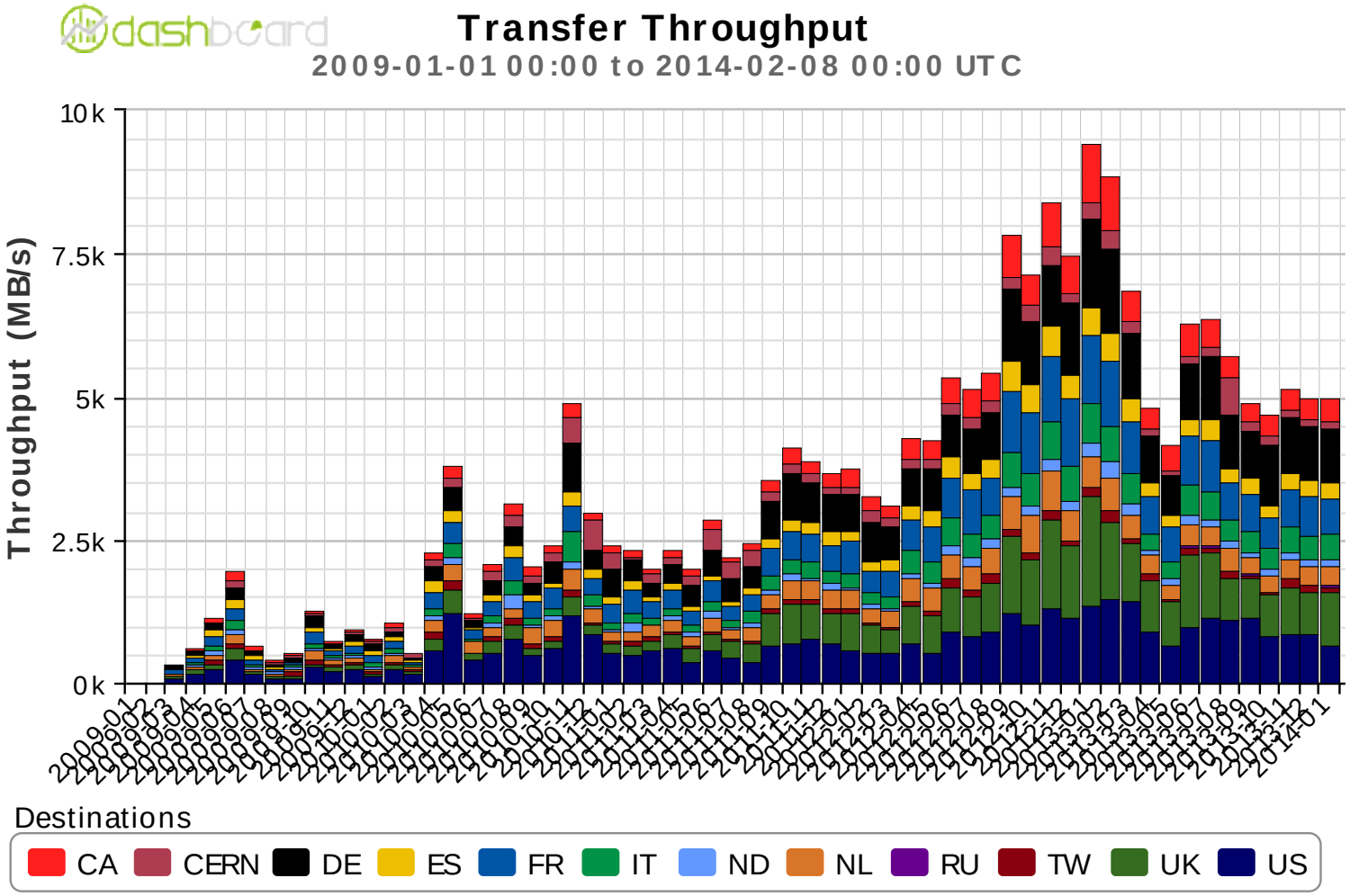
- We already spend more on disks than CPU

- "Solution": Rely more on the network for just-in-time or real-time data distribution.

FAX: Federated ATLAS Xrootd

Access any ATLAS data by name from T3, T2, T1 without first copying the data

ATLAS DDM Throughput 2009 to Now



Concerns

- Annecdotal evidence of throughput limitations
- 10 Gb/s becoming marginal for large T2s
- Not just the network:
  - Local network
  - File-server hardware
- Increasing reliance on rapid replication

- Global cost-optimization needed:
  - CPU recalculate or store
  - Disk/Tape store or recalculate
  - Network transfer or store

Implications for the Network

- Massive, policy-driven, predictable data distribution will continue, but growth will be modest.

- Bursty traffic (there are idle CPUs in xxx so replicate some data from yyy as quickly as possible) will become very important.

- Real time remote access to data will become important:
  - ATLAS does not yet fully understand how network bandwidth and latency will constrain this access
  - It won't be used where it doesn't work well!