



# Tape Planning

Richard P Mount



Tape Planning

Richard P Mount

Facilities Jamboree

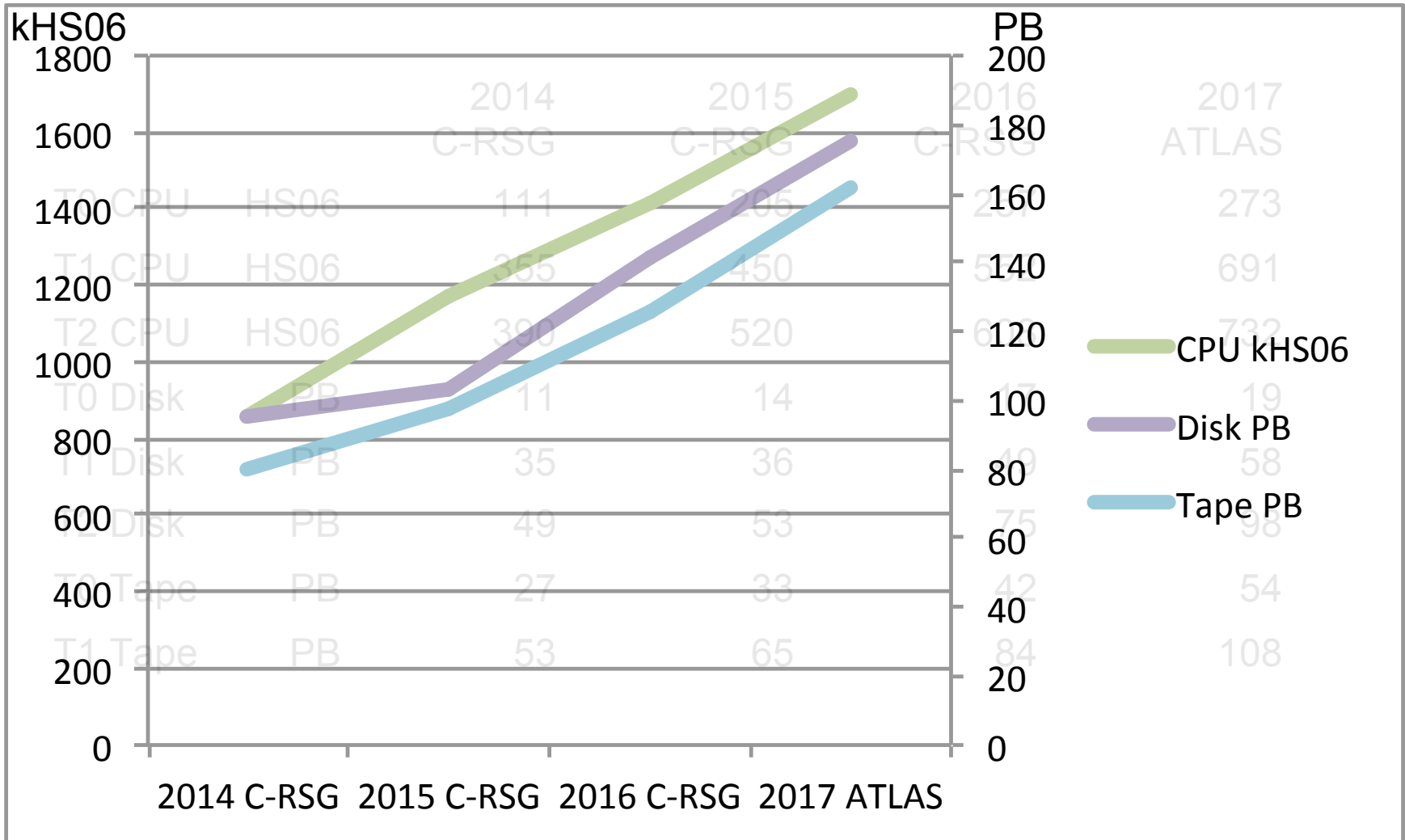
December, 2014



# Current State of C-RSG Recommendations and 2017 ATLAS Planning

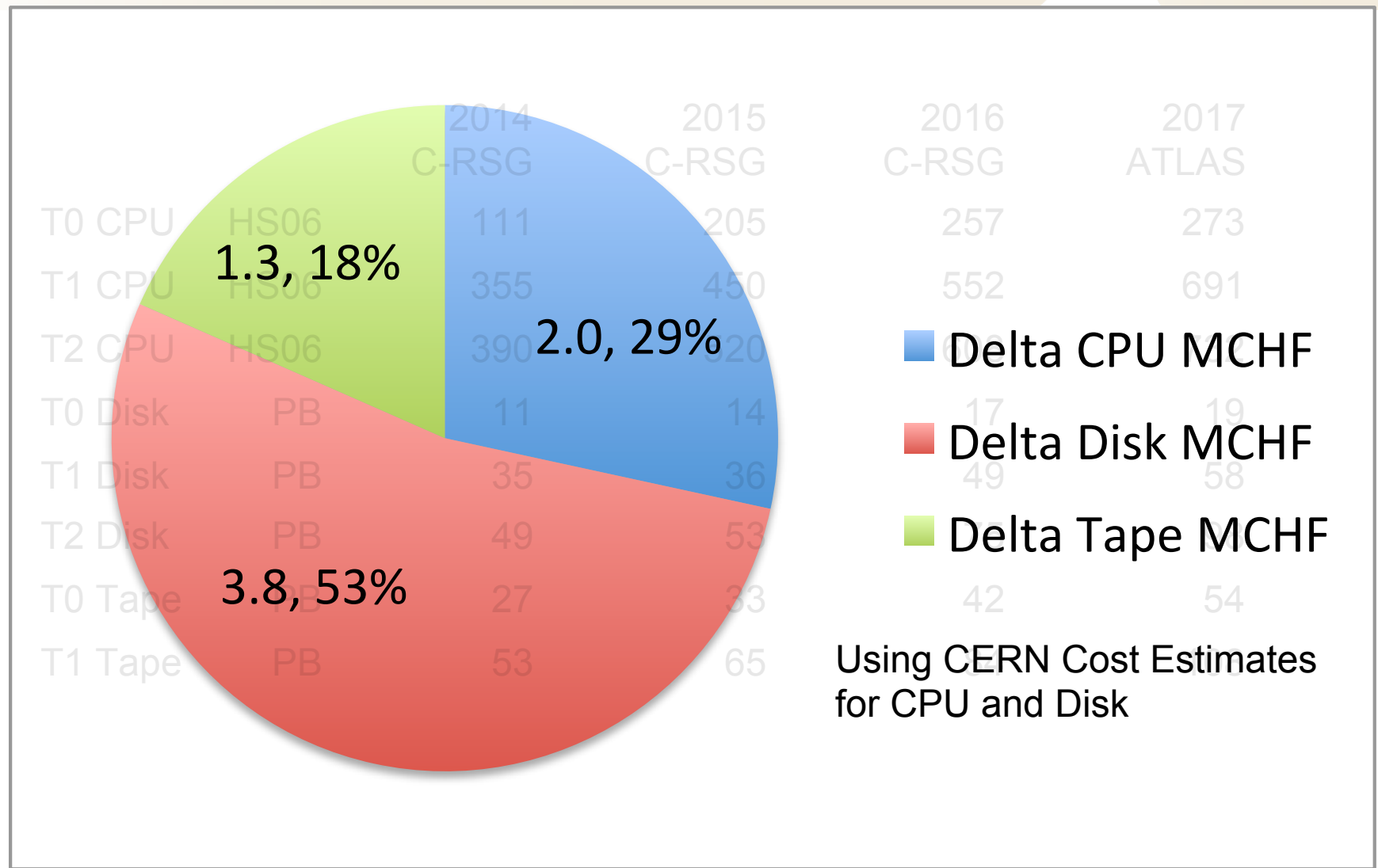


SLAC

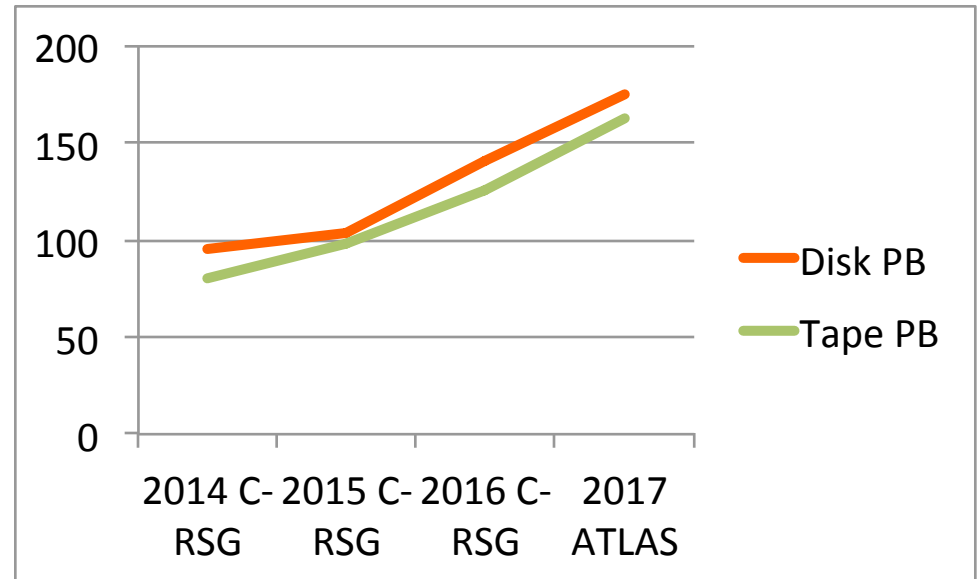
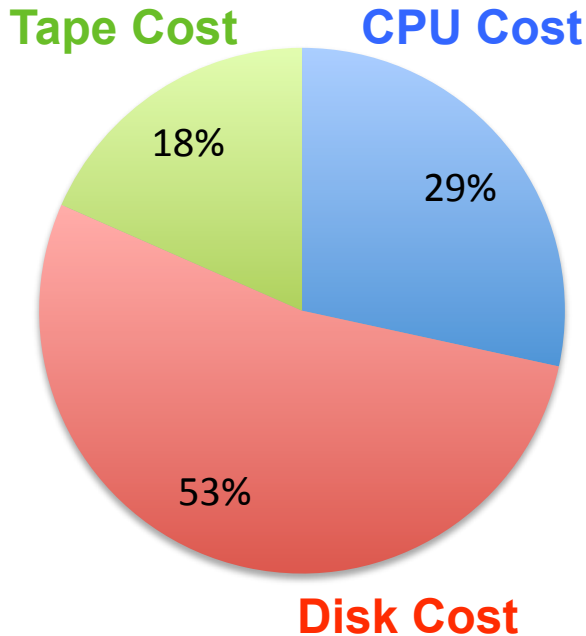




# Resources: The Cost of 2015, 2016, 2017 growth



# Tape Issue: Why do we make so little use of tape?



Tape is:

- “On its way out” (almost certainly untrue given dismal disk futures)
- “Inaccessible”
  - Tape latency (technical)  $\ll$  1 hour for small retrievals
  - Tape latency (for major retrievals): few days for human approval

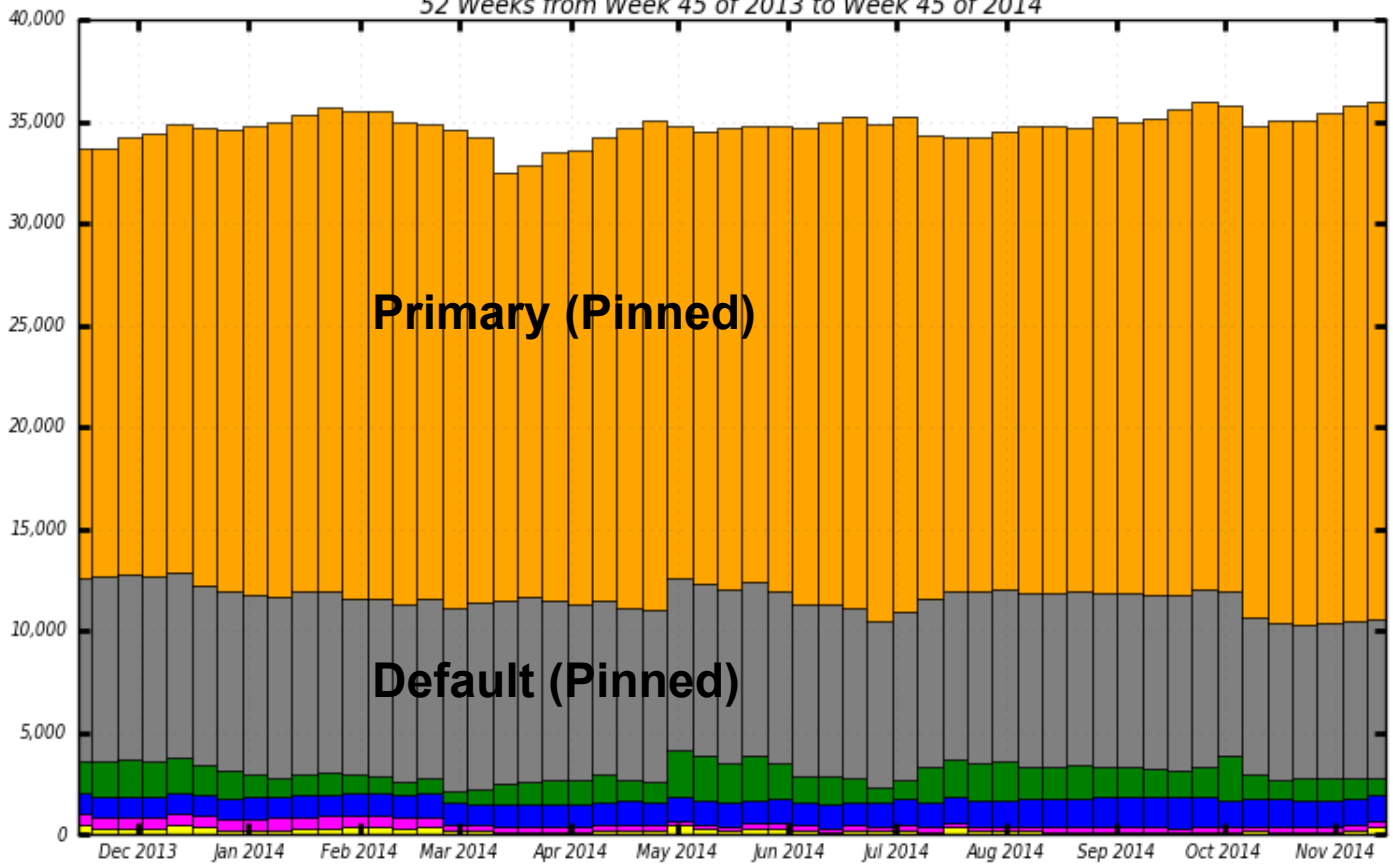


SLAC

# Disk Issue – 1: “Ongoing Tier-1 Disk Space Crisis



Number of Physical Bytes (in TBs)  
52 Weeks from Week 45 of 2013 to Week 45 of 2014



## Tier-1 Disk Space over last 12 months

“Secondary” data (The only data that can be deleted to make room for new data)



primary    default    secondary    input    extra  
tobedeleted    custodial

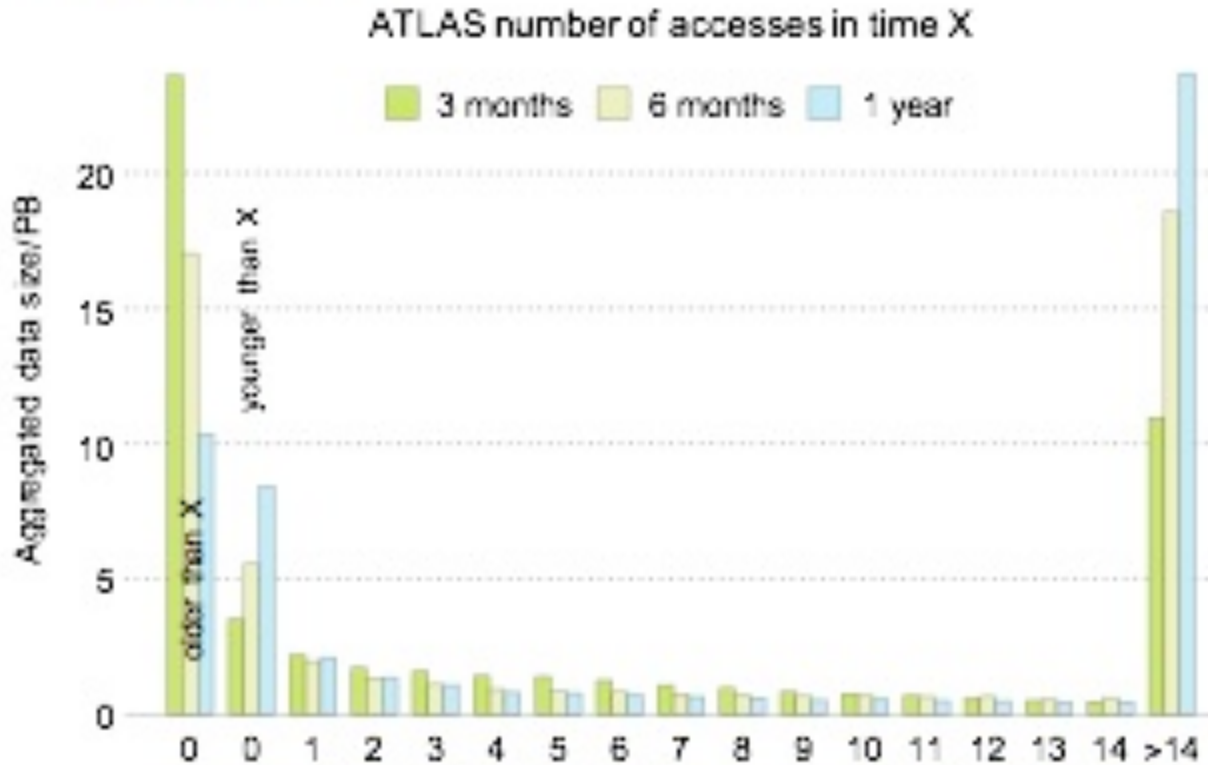
Maximum: 35,994 , Minimum: 0.00 , Average: 34,102 , Current: 35,994

# Disk Issue – 2: Unused Datasets on Disk (weakens the case for more disk)



SLAC

## Data popularity



Report from C-RSG  
to RRB

C-RSG = Computing  
Resource Scrutiny  
Group

RRB = Resources  
Review Board

Volume of data versus number of accesses in ATLAS DATADISK  
at T1s and T2s for 3, 6 and 12 months to 31 July 2014

14 October 2014

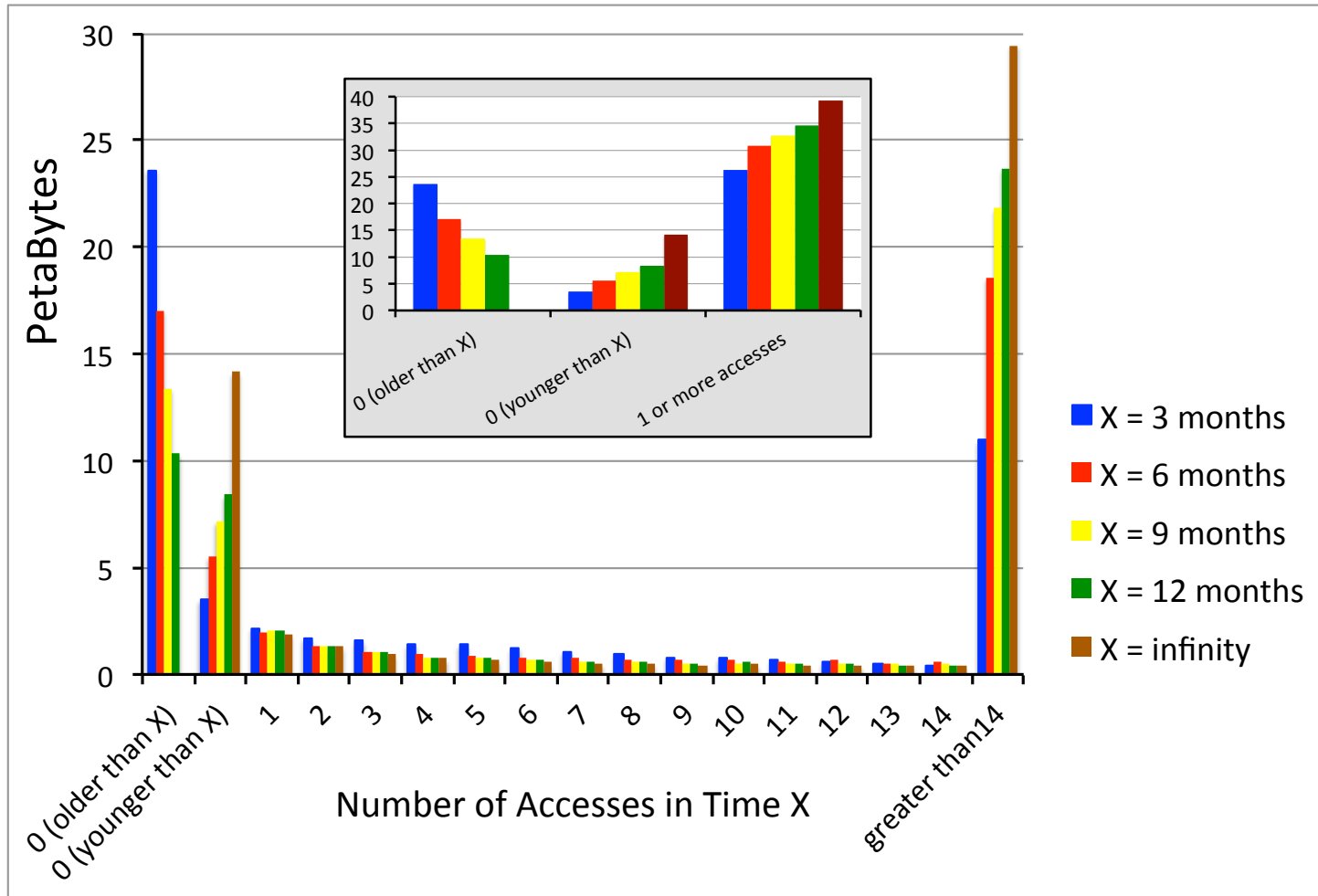
CERN-RRB-2014-080

12/27

# New Plot sent to the C-RSG



ATLAS DATADISK





# ATLAS Data Management Strategy - *Lifetime*

Assign a LIFETIME to all new and existing ATLAS data:

- ◆ Infinite for RAW data
- ◆ Likely ~2 years by default, but extendable
- ◆ Data older than the Lifetime will be DELETED from tape (and disk) with advance notification.

Principal consequence – clean up of tape storage

*plus ça change ...*

The planning spreadsheet I inherited from Borut had implicit draconian disk lifetimes

- No Run 1 data (or simulation) on disk after December 2015
- Typically 25% of “group data” (DxAODs and NTUPs) retained from year  $n-1$

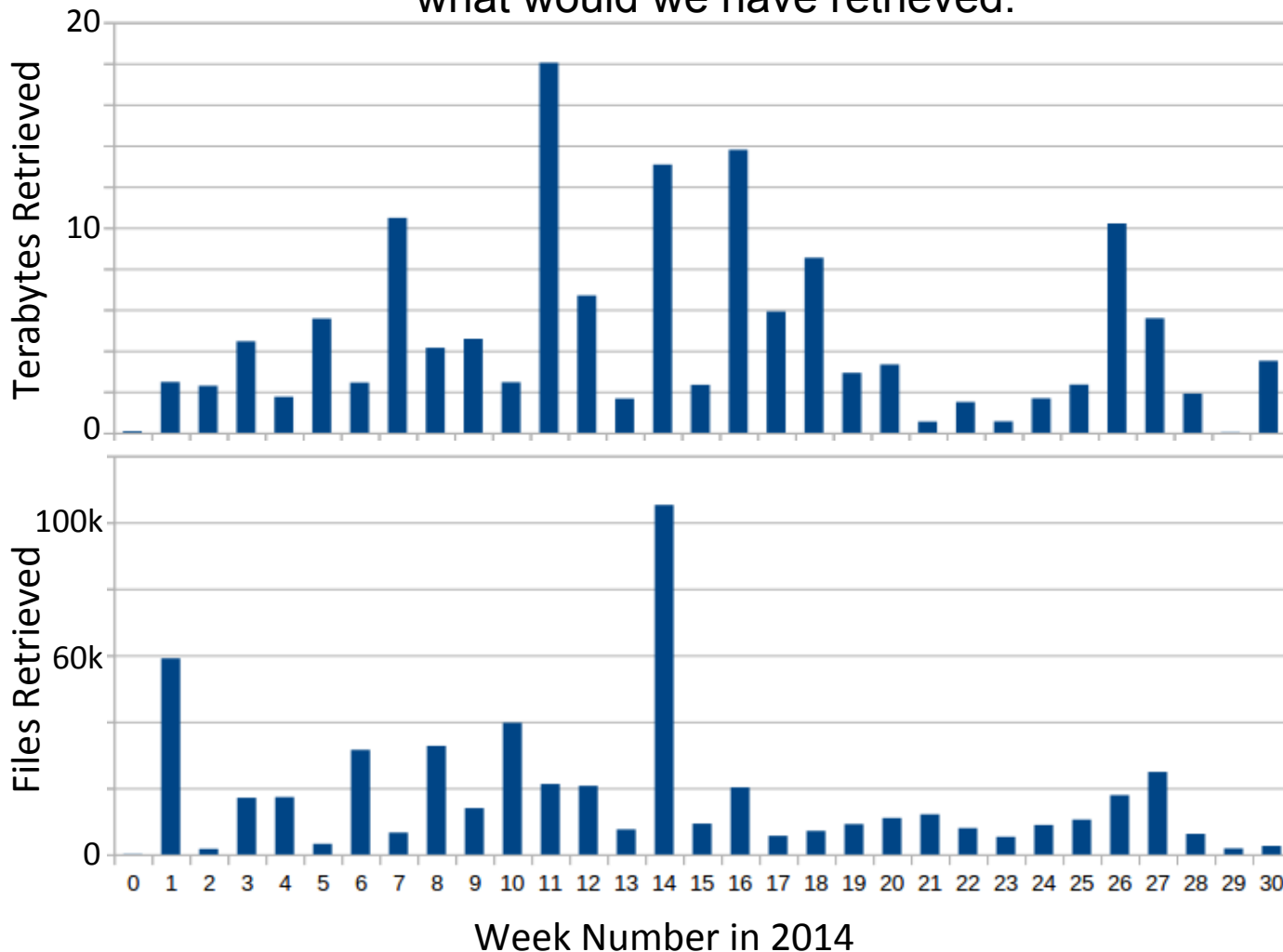




# Modeling the Impact – Access to Tape-Resident Data

SLAC

If all data not touched for 9 months (about 13 PB) were on tape on Jan 1, 2014, what would we have retrieved.



2% addition to “normal” retrievals

10% addition to “normal” retrievals



## Multi-Experiment Tier-1 Experience

“Our ATLAS tape system has an order of magnitude less activity than our XXXX tape system”

“Current planning is to reduce the number of tape drives serving ATLAS – they don’t appear to be needed”



## Tape growth thought experiment

- Increase tape system growth by factor 2
  - Decrease disk system growth by 33%
- Approximately unchanged cost
- Rate of increase of STORAGE multiplied by 1.56

Is this the optimal strategy?

Modeling + experience will tell us.