TEVATRON DATA CURATION

Gene Oleynik, Fermilab

Introduction

A little about myself

- Background HEP, Fermilab Fixed Target
- Joined Computing Department in 1987, Data Acquisition
- Working with storage systems at Fermilab since 2004
- Department Head, Data Movement & Storage

Outline

Focus on facilities for Run II bit preservation, mostly tape

- Storage Facilities For the Tevatron
- Technology Migration
- Environment
- Data Integrity, Monitoring and DR
- Moving Forward, Costs, Open questions and Risks
- Conclusions

Current Tevatron Storage

Tevatron Data located in two Data Centers on site

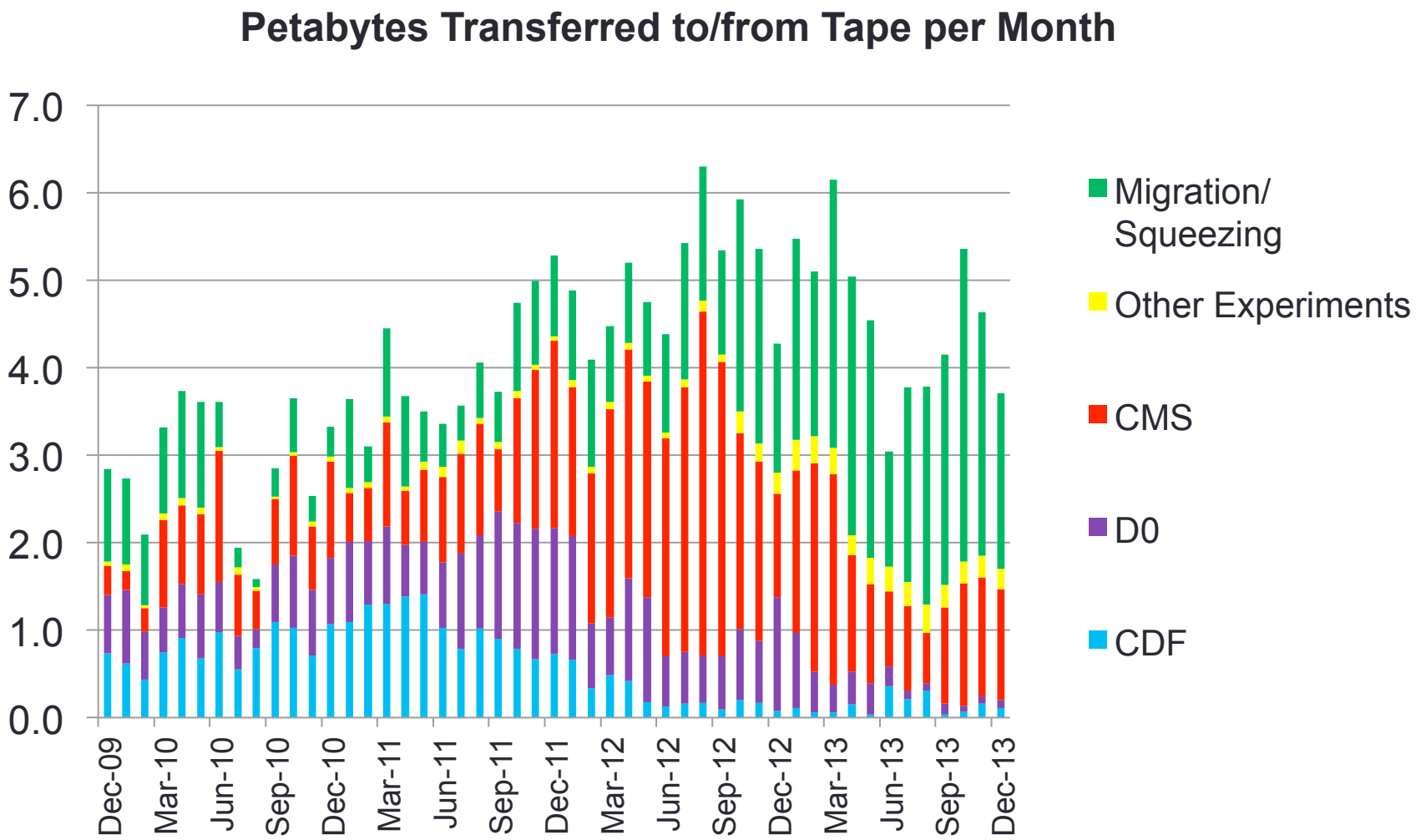
21 PB of tape storage about equally split between D0 & CDF
Combined front-end disk cache on the order of 3 PB

RAW and Online Database backups @ Grid Computing Center:
Tevatron uses 1 of 4 10000 slot libraries

Reconstructed and other data @ Feynman Computing Center:
3 10,000 slot SL8500 shared with other experiments
cache disks located here

Commitment to keep capabilities and data accessible to 2020

Current Tevatron Storage



Technology migration

Over the course of the Tevatron (not even counting 8mm):

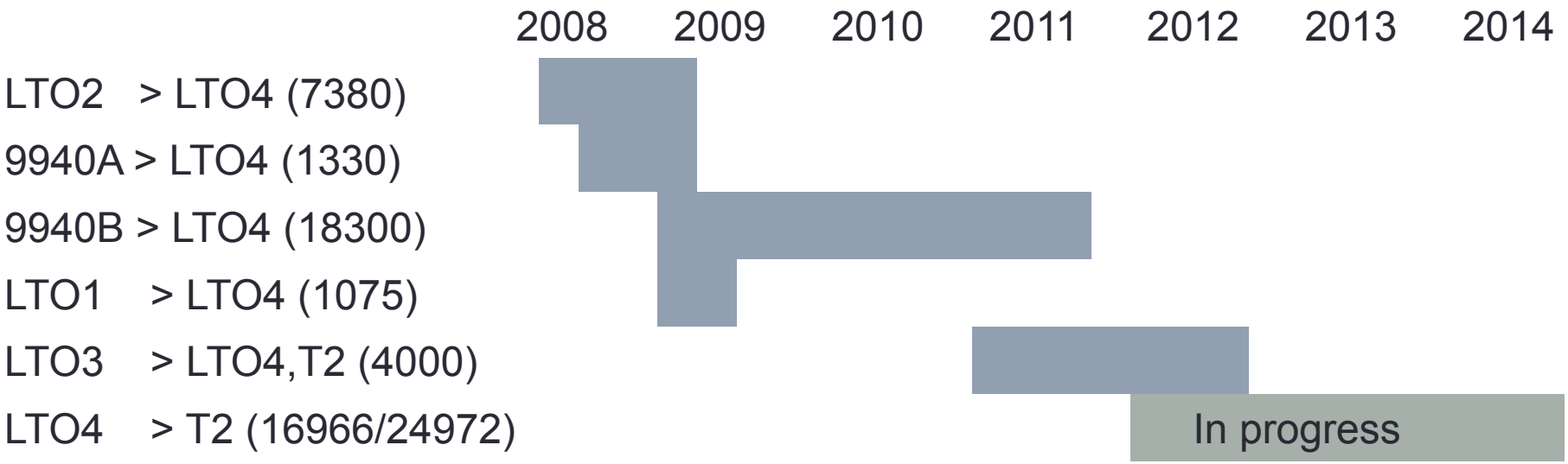
- > 90x increase in tape capacity
- > 24x increase in transfer rate
- Decommissioned 9310 & ADIC AML-2 tape libraries.
- Migrated off 9940A, 9940B, LTO1, LTO2, LTO3 to LTO4

Migrating LTO4 to T2 (5.4TB/cartridge media). 88% done

Care taken to insure all migrated data is copied and correct:

- Read back and verify checksum for every migrated file
- Validate metadata is correct
- Verify no file left behind when disposing of older media (new extra paranoid step)
- Ramping up migration took to a lot of effort and time. Use up to 8 “Migration Stations” in parallel.

Technology Migration



Parenthetical numbers are number of tapes migrated.

Migration Activities – Obviously a continual process

By the end of FY14, we will have migrated over 57,000 Tevatron media with a final count of less than 4000!

Technology Migration

Non-production commissioning of new technology. We scored a fail here

We missed a number of issues.

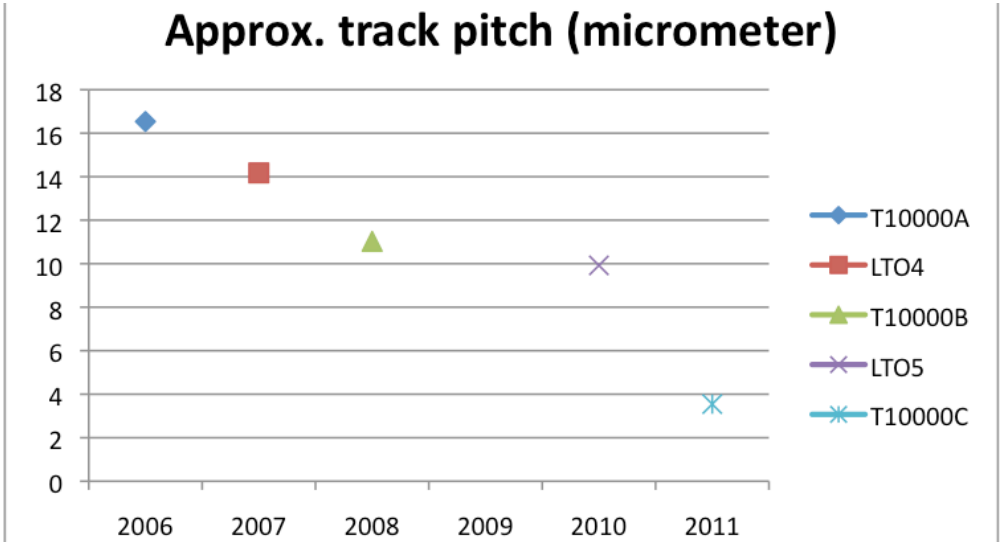
We now test in production by writing a primary copy in the old technology and a secondary in the new.

Continue to use in-house developed and HEP collaborative storage software:

- Enstore
- SAM, SAM cache
- dCache

Environment

Decreasing track pitch yields higher demands on the environment



Dimensional stability: Temperature and humidity results in creep and can cause read problems

Dust and other debris. Fine dust is of most concern.

Environment

We try to stay within the recommended operating range for Humidity and temperature:

	Temperature	RH (%)
Optimal Op	22C (72F)	45
Recommended Op	20-25C (68-77F)	40-50
Full Op Extremes	16-32C (60-90F)	20-80

For Dust, the recommendation is Class 8 cleanroom. We are about class 9 (better).

We still see dust built up in the libraries over the years, though it has not caused problems yet.

Environment

Q: What is the size of a Mosquito?

A: Around 20 (lost) files

We have had some small insect issues just recently in the GCC TRR.

Integrity, Monitoring and Disaster Recovery

Data integrity:

- End-to-end checksumming; Spot sampling files' checksums
- Experiment accesses (very good coverage while running)
- Write-protect filled tapes
- Extensive proactive health monitoring (soft errors, rates, etc.).

Environment:

- Wireless temperature/humidity recorders at libraries
- Portable industrial dust detector. Sampled around the rooms

DR:

- Data is mostly single copy, but CDF RAW is basically included in RECO, and The RAW and RECO are in separate data centers
- Online database backups at different data center than databases
- Second copy efforts started for CDF (FNAL to CNAF)

Integrity Issues We Have Encountered

- Fine debris buildup in LTO4 drives (bad tape batch?) resulting in slow transfers (like an hour to **successfully!** read a GB file), No data loss, required close monitoring to proactively replace drives. .
- Slitting Debris in a batch of T2 media. No data loss.
- Several instances of insects on tape. Some data loss (CMS T2).
- Mangled tape (very infrequent, though we just recently had such an incident with a CMS T2 media).
- Other firmware bugs – potential data losses (had copies).
- A number of unreadable files: 13/15M for Tevatron. We have never encountered a checksum error on tape, just sense media errors. Successful reading of files may be sufficient.

Integrity and Monitoring open questions

We currently sample randomly selected tapes and files and tapes and verify checksums.

- Is this the right thing to do in data preservation mode? ***Do we risk mangling a tape or catch a bug (literally) sampling the data?***

How do we measure Data loss?

- The real impact is lost statistical significance, and that varies (a calibration file vs. a RAW data file).
- Easiest to do is lost files or potentially lost (is it lost if it exists elsewhere?).
- A work in progress.

So what does all this cost?

Amortized costs (M&S costs over the appropriate lifetimes)

- Tape and disk hardware
- Infrastructure equipment, servers, network switches.
- Migration (media amortization ~ 6 year), tape trade-in, decrease in tape cost over time

Yearly costs:

- Salaries: 5.5 system admin, 4.5 developers
- Facilities (electric, building)
- Maintenance

Lab overhead costs for staff and M&S

Duty factor: assume 50% of library occupied by customers

Estimate ~ $25-30/TB/yr for tape 5-10x this for disk

Moving Forward

Minimize the differences in technologies by the experiments and support sustainable ones.

- Plan to stay on T2 media for some time. Complete migration to T2 by the end of FY14.
- D0 plans to move to SAM+dCache rather than SAM Cache

Reduce the amount of equipment to support

- CDF plans to reduce their cache disk from the 2011 level down to about 6% of that by 2016. D0 will likely do similar
- T10000C tape drive count has been reduced from LTO4

Fermi >> CNAF copy for DR (Silvia talk)

What keeps me awake

Large unplanned effort and costs:

Is the software technology sustainable. In-house and collaborative expertise sustainable?

- This can be expensive to move from: new interfaces, formats, data migration and etc. may all be needed

Reliance on proprietary vendor technology

- We use widely adopted hardware, but it could be costly and require a costly migration to different technology if there is a vendor issue

Dust Cleanup?

Conclusion

We are on track on our commitment to maintain Tevatron data through 2020.

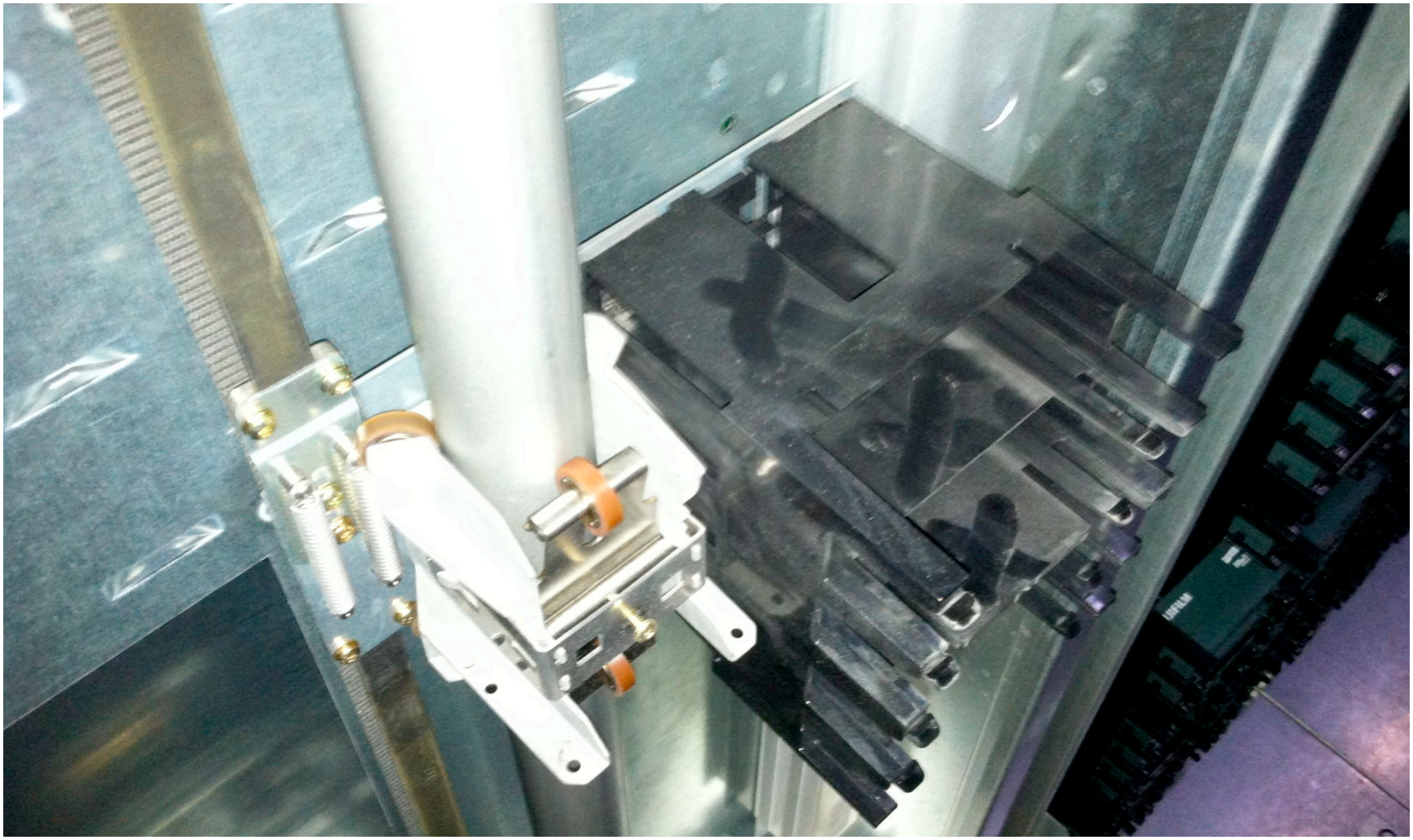
We have had a number of bumps but have worked through them with little impact to the experiments.

Questions?

Backup slides

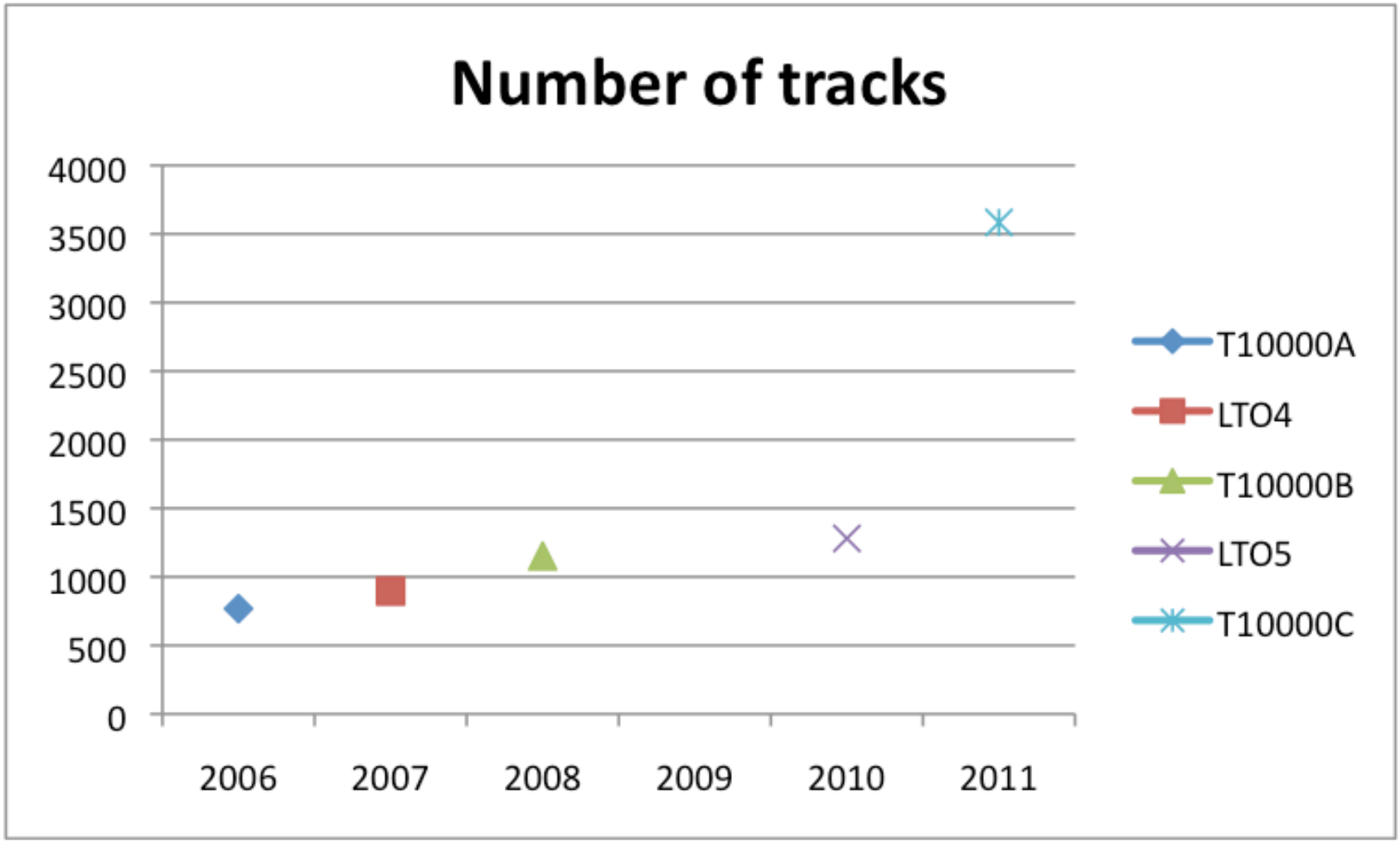
Dust

Dust buildup in one location in a library

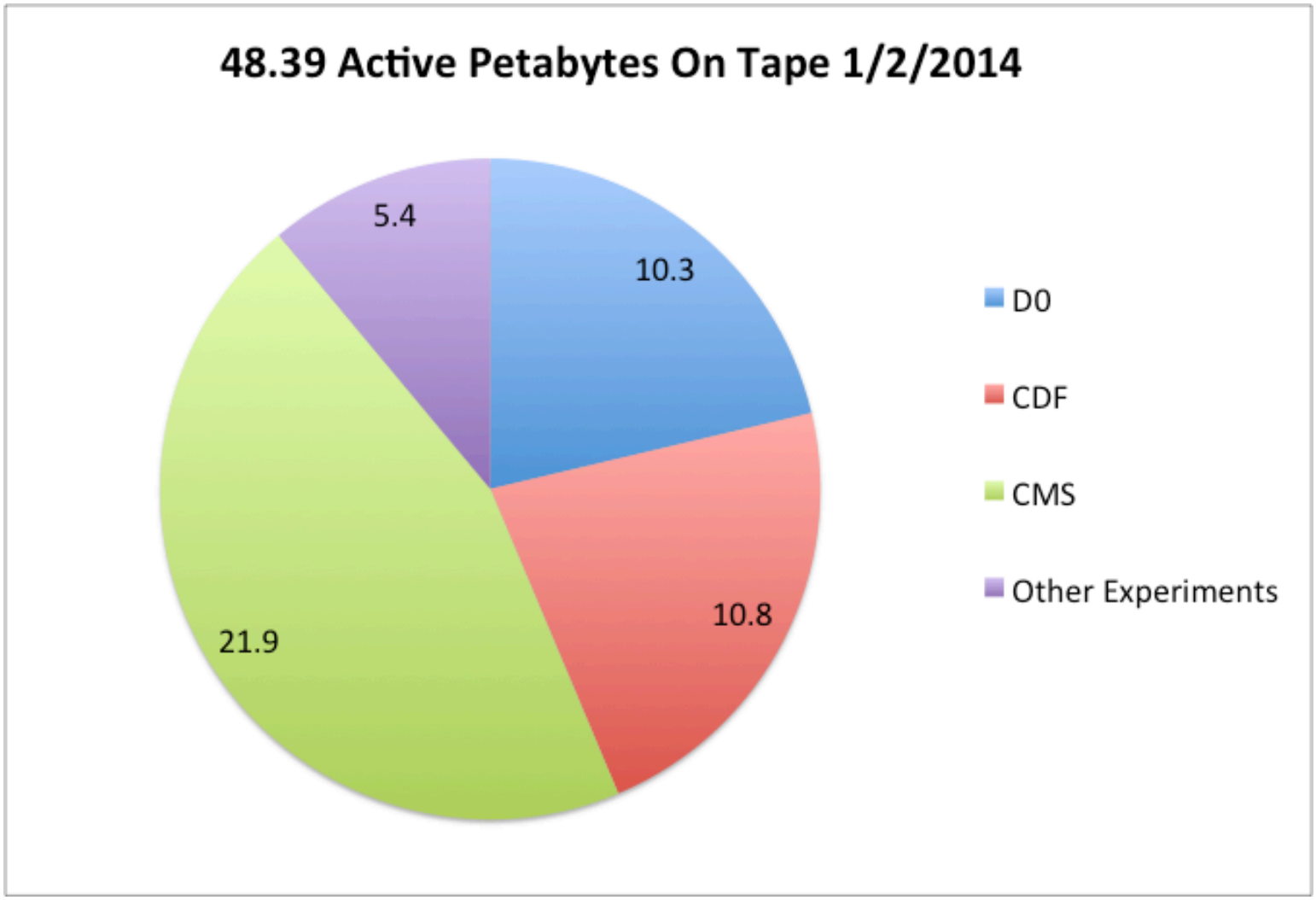


Snake

Tape track density trend



Distribution of Active Data



Media distribution at FCC Libraries