LC STORAGE ENVIRONMENT (FY’16 – UPDATE)

Library of Congress
OCIO/ESE/SysDev
Carl Watts
Digital content continues to grow
- 525TB (61+M unique files) added to online access
  - Over the past 12 months Online growth was 48%
  - FY’15 growth was 34%
- 2.9PB (31M+ unique files) added to long-term storage
  - FY’16 was 26%
  - FY’15 growth was 50% primarily due to data migrations from older systems
Storage Tier (*Available Capacity*)

- **Tier 0**: 255TB
- **Tier 1**: 820TB, 437TB
- **Tier 2**: 11PB, 9.3PB
- **Tier 2b**: 11+PB, 9.3PB
- **Tier 3**: IBM GPFS-LTFSEE, 11+PB, Oracle OHSM, 66+PB
- **Tier 4**: IBM TS1140 Tape, 560TB, Back-up, 16+PB, Oracle T1000D Tape

- **Legacy**
- **Dynamic**
- **Hybrid**
- **Static and Cache**
- **Disk-to-Disk**
- **Tape**
Still working on migrations…

**OHSM tape migration from T10KB to T10KD**
- Copy 1 took 10 months longer than the original estimate
- Copy 2 turned into a migration of a migration, still ongoing

**Migrating from our HPSS to LTFS EE is ongoing**
- Current planned completion 6/2017

**Data consolidation continues**
- Four (4) presentation server migrations in FY’17
We tested 5+ object storage system in 2016

Use cases tested

- Presentation storage
- Archive storage
- Object to POSIX via NAS gateway and s3fs
- POSIX to Object via GPFS
Object Storage Review

- Observations
  - Not all Object Storage systems are equal
    - Five systems tested, five different results
  - There are no standards (S3, OpenStack, CDMI)
    - Compatibility is not a standard
  - Weigh the environment choices
    - Consistency vs. Eventual Consistency
  - How do you plan to use the system?
    - DevOps vs. Curatorial
  - Sizing the environment
    - Memory is your friend
  - Some middleware is needed
    - Maintaining abstraction layer for POSIX and Objects
email:
cwat at loc dot gov