Library of Congress' National Audio Visual Conservation Center

Update 2019

Scott Rife
Chief of Data Center Management
IT Services Operations

The Packard Campus

Mission

- The National Audiovisual Conservation Center develops, preserves and provides broad access to a comprehensive and valued collection of the world's audiovisual heritage for the benefit of Congress and the nation's citizens.

Goals

Collect, Preserve, Provide Access to Knowledge

- The National Audiovisual Conservation Center (NAVCC) of the Library of Congress will be the first centralized facility in America especially planned and designed for the acquisition, cataloging, storage and preservation of the nation's collection of moving images and recorded sounds. This collaborative initiative is the result of a unique partnership between the Packard Humanities Institute, the United States Congress, the Library of Congress and the Architect of the Capitol.
- The NAVCC consolidated collections stored in four states and the District of Columbia. The facility boasts more than 1.5 million film and video items and 3.5 million sound recordings, providing endless opportunities to peruse the sights and sounds of American creativity.

Packard Campus – Many formats

NAVCC Technical Status

- Current: 10.6 PB and 2.4 Million files replicated in 2 locations.
  - Compared to 7 PB and 536 Million files for Newspapers, internet archive, prints and photographs, etc. Long Term Storage (LCBP)
- 50+ Points of Digitization (PODs):
  - Podcast serials programmatic ingest has started
  - 4K preservation for film is growing
  - Programmatic Orderless ingests for Senate, NFL, and Saturday Night Live
  - 34 Solos (16 in robotic cabinets), 9 Pyramix, 10 Linux(OpenCube, etc) 1 Quadriga, 2 DVD Rippers, 1 CD Ripper, Oxberry, Arrilaser, Spirit, Vario, Clipster
  - Daily each POD can generate: 2GB-150GB for audio and 50GB-1,200GB for video
- 415,000 square feet, more than 90 miles of shelving for collections storage
  - 35 climate controlled vaults for sound recording, film, and videotape
  - 124 individual vaults for more flammable nitrate film

NAVCC Business Status

- Video Preservation Master: MKV files (FFv1 in a Matroska wrapper). NAVCC is consulting with FADGI to begin using in addition to MXF files (JPEG2000). FADGI will conduct research to determine tech specifications. March 2020 expected to finish. AAPB has asked the Library to accept FFv1/Matroska as a result of NEH digitization grant.
- 4K scanning increasing: Ensure digital preservation of original film elements to capture full resolution of each frame from the following formats: Camera original elements including negatives for both 35mm and 16mm, Reversal positives (Kodachrome), Fine grain masters (made photochemically from camera negatives)
- Size matters: 4K files require specialized hardware to playback, Ingest rates are affected by larger files, Failure to ingest after 1 TB file moved over hours means lost throughput, Amount of data processed each day increasing, Security redesign reduced throughput
- NFS mounted OHSM filesystems required tweaking (DIO disabled)
- 2K scanning covers the rest (anything more than 2 generations away from camera original typically only hold resolutions to 2K)
- Rewrite of Packard Campus Workflow Automation (PCWA) provides a better platform for Agile development, but resulted in short term pain as new application is rolled out
- LC will be releasing an RFP for a collection management system for audio-visual content (AVCMS). This effort will replace the existing MAVIS software and some functionalities of the PCWA platform. Digital collection content will not be stored within AVCMS, but will remain in the custody of LOC.

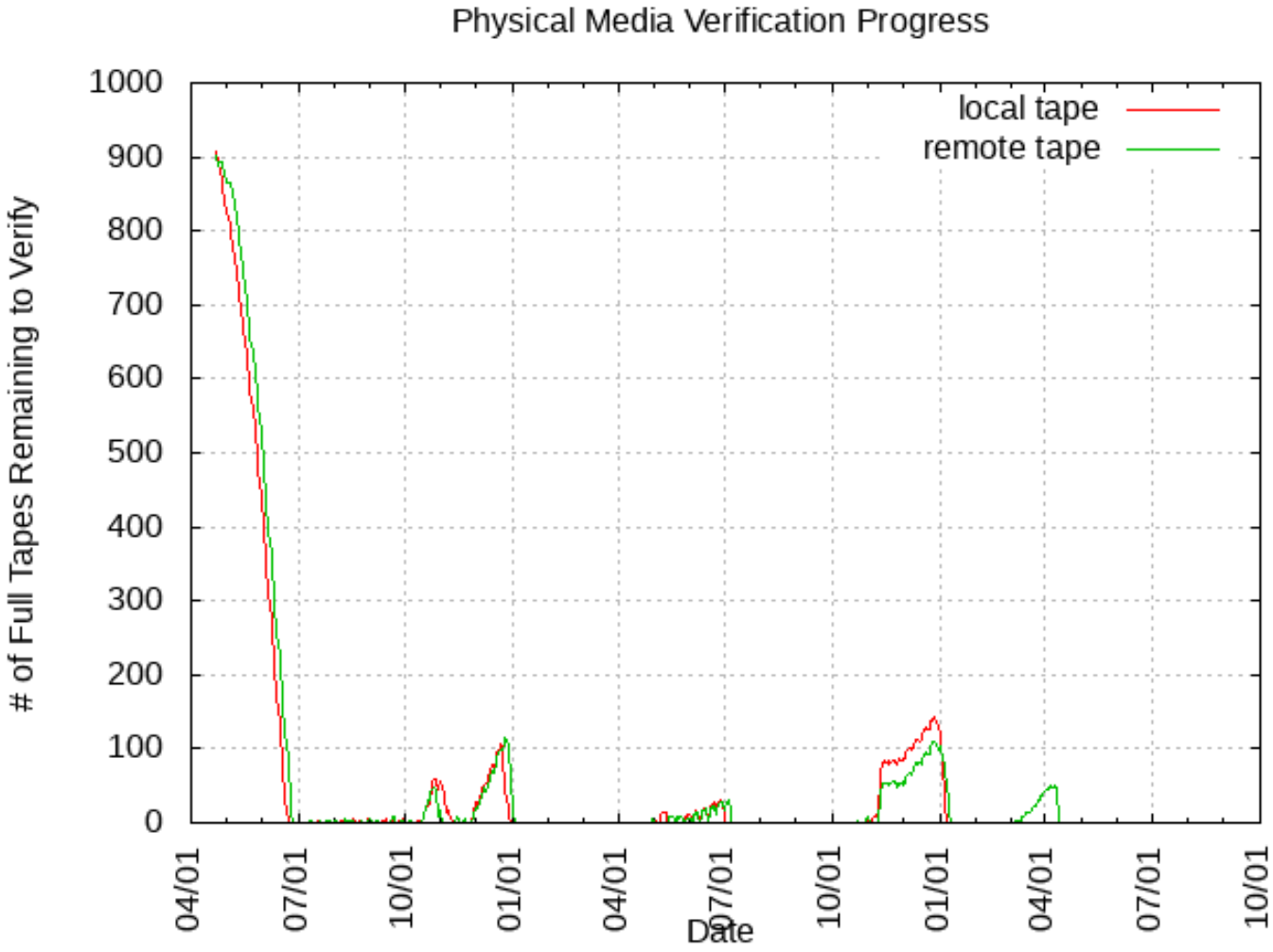
LOC Verify and Correct at the Marginal

Protecting content is different than data loss

- Reduce the likelihood of content loss while recognizing that data loss is statistically likely
- Catch and correct all marginal errors and failures as soon as possible
- Verify all the content at a regular interval
- Some of the regular verification processes that we run:
  - Samfsbackup (meta data backup) 5X/day
  - Verify samfsbackup size and frequency. Send an email if missing.
  - Fix damaged files. Occasionally a file will be marked damaged because it cannot be retrieved from tape. Usually because a tape was stuck in a drive/robot/pass thru port. Find these everyday and attempt to stage. If we can't, then send an email. Send an email when we find damaged files so we know issues are occurring and being corrected
  - Stats: Watch the # and size of files waiting to archive. Warn when the # of files or size of files exceeds thresholds. Usually an indication of some marginal error condition. Fix before file system fills up or we fail to deliver a file for customers.
  - Samfsck: Run this daily with filesystem mounted. Warns when there are marginal conditions with file system before they are catastrophic.
  - # of tapes/TB available: Know when we are running low so we can correct before a failure
  - Tpverify: Verify all tapes with data every 6 months. Verifying header and all blocks of data on tape with CRC.

LOC Media Verification (Content and Tapes)

- Finishing full verification of 10 PB / 2.4 M files compared to original fixity generated before ingest.
- May to September 2019
- Enhanced algorithm separates staging from fixity calculations
- Allows for recycling of over 50 pallets of physical media



NAVCC Observations

- Next migration will be LTO, if tape is still the best TCO
  - LOC is putting a copy in the Cloud where effective.
- SSD offers power and cooling that fits NAVCC limitations
- Customers require TB of short term storage for projects
  - Force them to request for short periods. Program deletions from shared NAS
  - Provide embargo space for longer term, inexpensive storage