Library of Congress' National Audio Visual Conservation Center

Update 2018

Scott Rife
Information Technology Specialist
IT Services Operations / Operations and Maintenance / Unix Systems

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 Current state

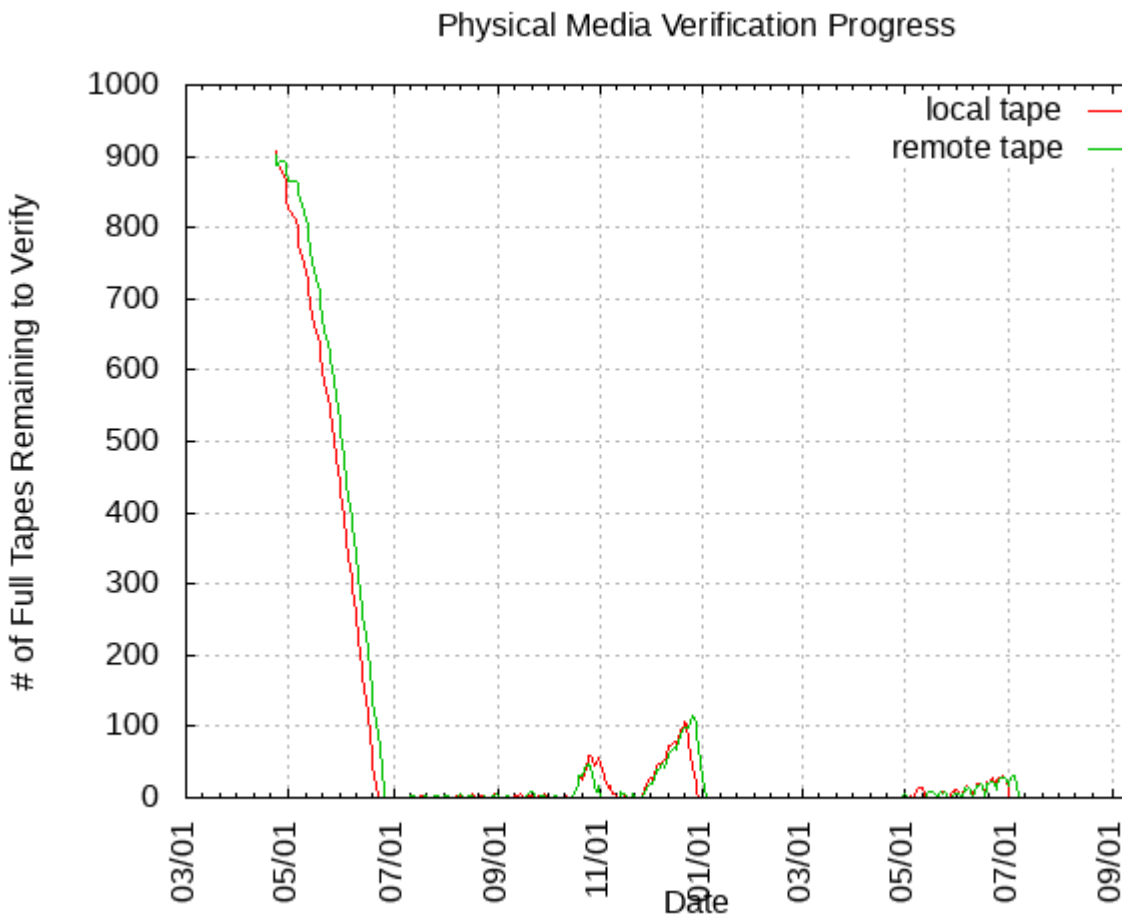
- Current: 9.6 PB and 2.2 Million files replicated in 2 locations.
  - Compared to 5.1 PB and 410 Million files for Newspapers, internet archive, prints and photographs, etc. Long Term Storage (LCBP)
- 53 Points of Digitization (PODs):
  - 4K preservation for film is new this year
  - Programmatic Orderless 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
- Additional workflows coming in the future include Born Digital Copyright Audio submissions (Podcasts), Live capture-264 DVRs, PBS and others.
- 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

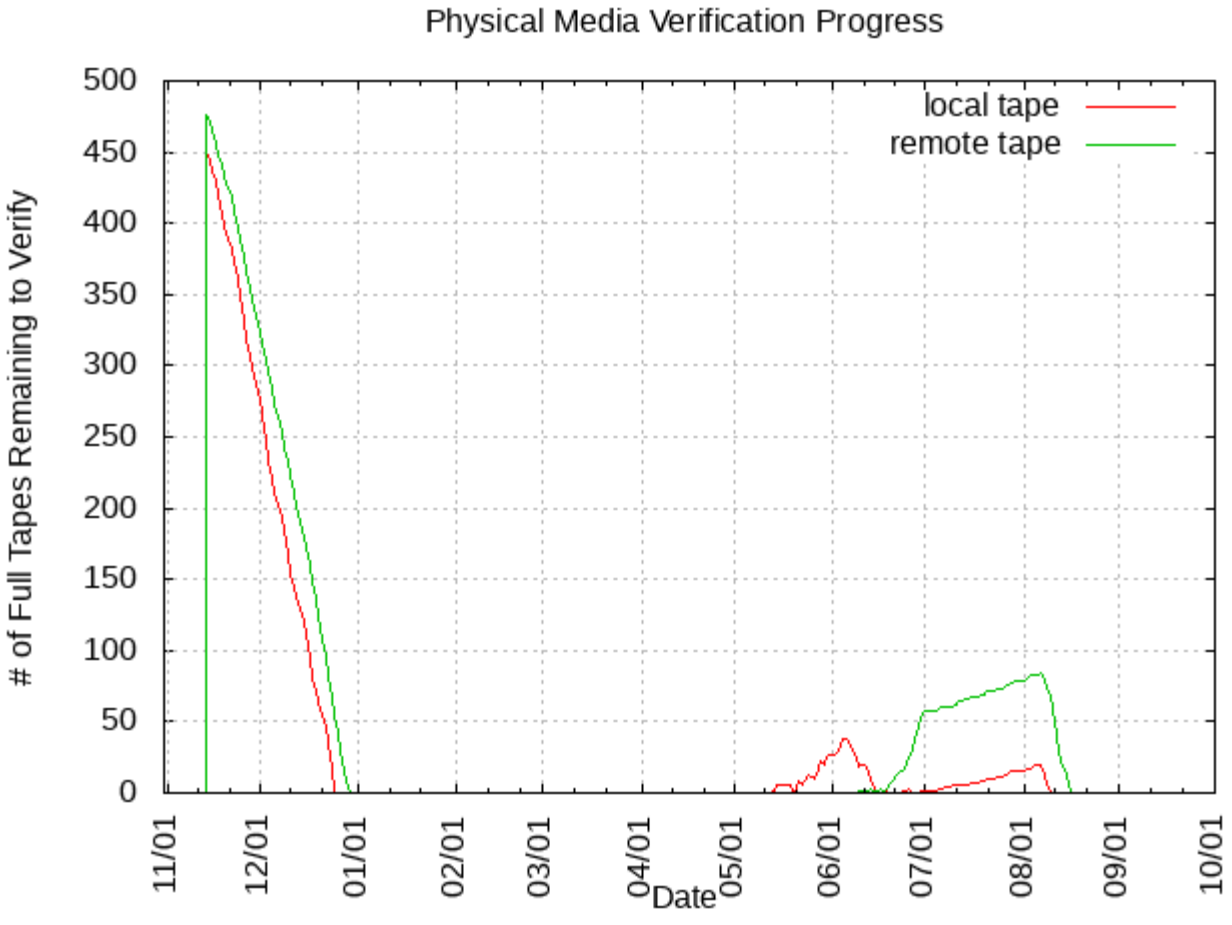
LOC Verify and Correct at the Marginal

Content is different than data

- Reduce the likelihood of content loss while recognizing that data loss is inevitable
- 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 (NAVCC and LCBP)

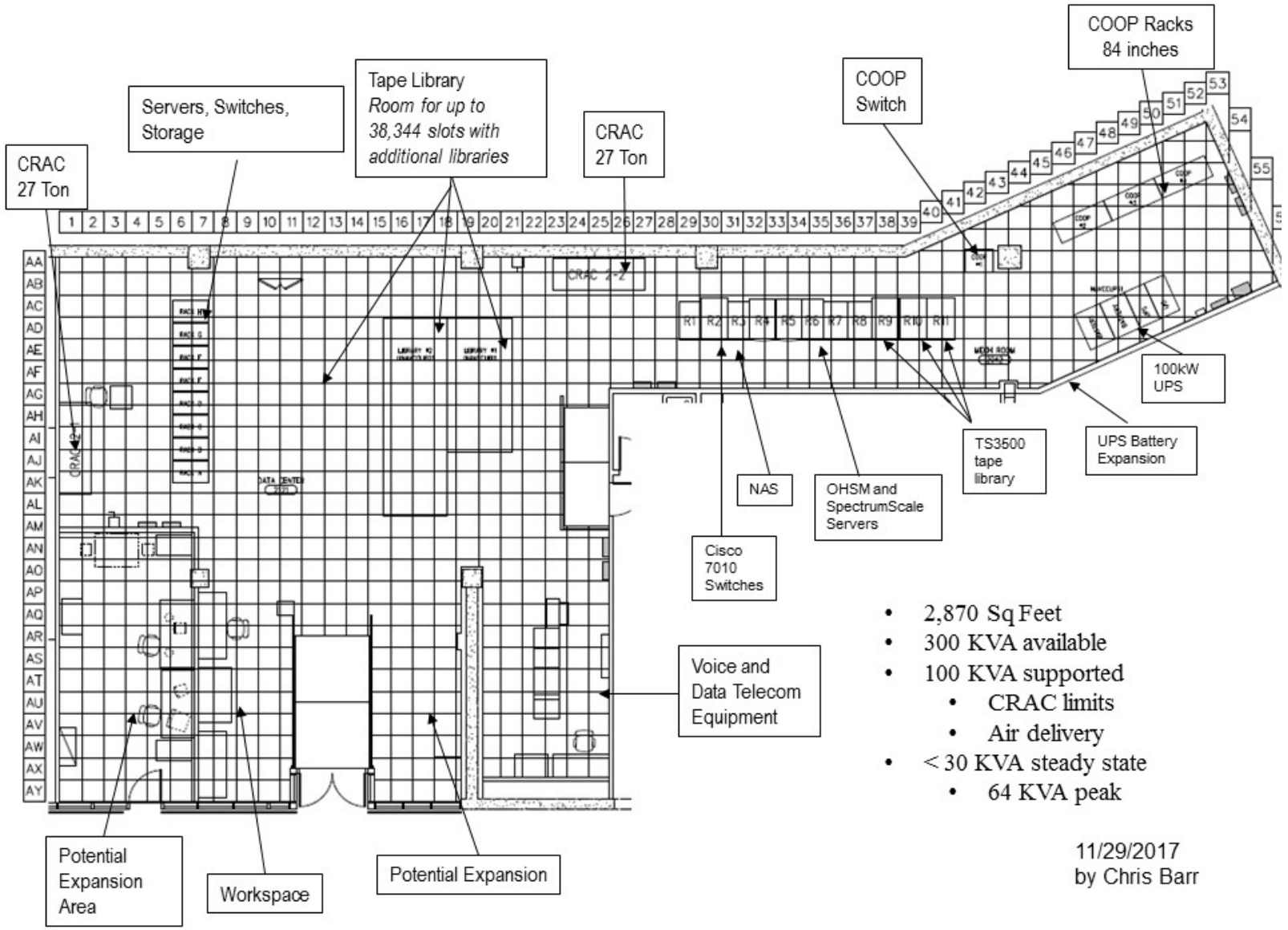




NAVCC Observations

- Migrating from T10K-C to T10K-D was cost-effective. Doubled the capacity and reduced floor space requirements for future growth.
  - Plan on reducing tape cabinet needs and clearing floor space
  - Last migration completed without issues
  - Verification after write to tape necessary, even if only sampling
- Next migration will be LTO, if tape is still the best TCO
  - LOC is taking advantage of Cloud where effective. Copy of last resort in Cloud?
- SSD offers power and cooling that fits NAVCC limitations:1 PB in 1 RU?
- Customers require TB of short term storage for projects
  - Force them to request for short periods. Program deletions from shared NAS

NAVCC Facility



- 2,870 Sq Feet
- 300 KVA available
- 100 KVA supported
  - CRAC limits
  - Air delivery
- < 30 KVA steady state
  - 64 KVA peak

11/29/2017
by Chris Barr