MIPoPS

Wednesday, December 4th, 2019

10:00-12:00, Seattle Municipal Archives, Conference Room 370

- Best practices isn't one size fits all!

Libby Hopfauf, MIPoPS & Seattle Municipal Archives
- Can't just send people on their way with hard drives, after failures we realized we needed to be doing and providing more instruction for digital preservation.
- Overview of AV preservation file formats
  - Different wrappers - AVI, MOV, MXF - many are proprietary.
  - Based on expert advice and recs., we use Matroska wrapper (.MKV) in 10 bit FFV1
  - Most people don't have space to store huge uncompressed files - dealing with limited space.
    - Lossless compression for video files is okay.
  - Provide LTO dark storage backup for our PIs files.

Sarah Shipley, Seattle Municipal Archives
- Developing a Digital Preservation Plan
  - What is it? Who, what, why, when, how.
  - Based on Trusted Digital Repository
  - Based on what youre doing now, and what you're trying to accomplish.
  - Makes a clear commitment for your organization
  - Addresses challenges, a tool for helping you get the resources you need to carry out your plan.
  - Objectives - what tasks need to be done?
  - Mandate - where does the mandate to preserve your records come from? Mission? Legal?
  - Scope - what are you preserving?
  - Operating principles - What standards are you adhering to?

- Roles & Responsibilities - whose responsible for doing what?
- Access & Use
- Challenges & Risks - These can set expectations for your org. Identifying the challenges, so you can argue for the resources that we need to do it…
- Audit & Transparency - How are you going about this?
- Framework administration - plan for long term stewardship.
- Sources for your policy!

Hannah Palin, University of Washington Libraries, Special Collections
- Sometimes a plan doesn't exist. Have to scavenge for equipment and make it up as you go along.
- Ground up - how do we approach this?
- Guided by getting the materials out to the public
- Very iterative process.

Ann Lally, University of Washington Libraries, Special Collections
- Recently redeveloped digital preservation processes after a [Lean](#) process
  - Students asking "why are we doing it this way?", pushes her to justify and re-think standard processes. New inputs can result in better processes.
  - Disk image created using [FTK Imager](#)
    - Review disk image in [BitCurator](#)
    - Not removing any personally identifiable information (PII), arrangement, description on ingest.
      - PII problem - if we create the list of PII now, and it isn't used for 50 years, will it even be translatable?
      - Instead, doing it as items are requested.
    - Not de-duping (not worth the storage) using [DeDuper](#)

Moriah Caruso, University of Washington Libraries

- Chance to think more broadly about digital preservation across the institution
- Developing the born digital processing pipeline - files themselves, and removable media.
- "Digital" is big and spread across the organization - necessitates robust policy and planning.
- Search "UW Digital Preservation Policy" - not day to day workflows, but high level. Required approval of administration.
- Happy to share scaffolding that underlies the program!
- Gaining intellectual control over what needs to be preserved: How much is there, how fast does it grow? What's been sitting around for 20 years? Magnitude.
- Archivematica is the pipeline into the libraries storage.
    - Identify file format, validate against PRONOM registry, packages, Bags are like a grocery bag and receipt.
    - Bagging as soon as they get it, to establish chain of custody from day one.
    - Bagger / Bag-it = Maintained by LOC, but is open source
    - Different ways you can use this tool that'll give you the receipt for your bag.
- Doing everything in Amazon Web Services.
    - Decided to go that way because we don't do PII, don't process digital materials prior to ingest, but UW has HIPPA & security agreements with Amazon, files in the cloud are secure.
- Contract with Artefactual to maintain their system (Archivematica)
- Not all unicorns and light! Tension between trusting vendors, trust but verify!
    - Monitoring storage environments.
- Often feels like building the airplane as you're flying through the air
- Creating handmade delivery process for DIPS
- Chronopolis - UCSD super computer center, as super backup
- NDSA matrix is excellent for starting out

MIP❂PS

[Marty Gengenbach, Gates Archive](#)

- Acquiring Content from Optical Media at Gates Archive
    - How Optical Media Works
        - Written on a plastic compressed disc as a big spiral from the inside out
        - Data is written as pits and lands. Laser is shone on those pits and lands
        - Translated by computer into binary
- Problems with optical media
    - Degradation - physical object falls apart over time. Physical damage to the bitstream because it's very vulnerable.
    - Logical Structure - Sessions containing multiple tracks. Each track may contain multiple files. Varied and complicated structure can make reading the disc difficult.
    - Variable File Systems - ISO 9660, defines the form and structure for data stored on optical media; but there are multiple file systems that may exist on the same disc.
- Acquisition workflow
    - Attempted conventional file pull; mount, copy the files, collect metadata.
        - [Preservica](#) used to package and preserve
        - [EAC (Exact Audio Copy)](#) - captures audio tracks and a TOC (table of contents file) with metadata. Media carrier image is also captured and included in preservation.
        - BitCurator - open source set of tools for aquriing and analzyingdigital content fromanalog media carriers
        - [CDRDAO](#) - extracts audio content, captures an entire data dump of the disc content. Want to accommodate the capture of as much data as possible.
        - [Guymager](#) - Forensic disc imager. Doesn't work a lot of the time

- Creates forensic disk images - allows you to put metadata and checksums for individual forensic disc image segments.
    - [DDrescue](#) - reads full track forward and backwards, skips over bad sectors and recovers bad sections.
    - Creates a map of bad sections
    - BitCurator disk image access tool - generate the actual file content and access copies.
- Prioritize getting data off the carrier, where it can be stored and managed
- Embraced the value of the partial capture
- Define technical success and failure - constant cost benefit analysis. How much time can I spend on this one disc when I have 200+ to work on?
- You don't always need to create a disc image. When is there something inherent in the object, and when is it truly just a storage medium. Curatorial and resource based decision.

Josh Zimmerman, Archives of the Catholic Archdiocese of Seattle
- Employee files; Digital photographs (generally around events)
    - Made decision not to ingest them into Archivematica
- What is Archivematica?
    - Microservices!
    - Bundle a bunch of digital preservation actions into one platform.
    - IT didn't like it because it is open source and "unsupported".
- Archives Direct - Professional Archivematica
    - Fully hosted web based service, with back end storage on Amazon Web Services
        - Didn't have to rely on IT for anything
- [QuickView Plus](#) - can view obsolete file formats
- Some problems with Archivematica
    - Issues downloading AIPS over 1 GB

MIPoPS

[Ari Lavigne, Moving Image Preservation of Puget Sound](#)

- VideoAIP and AudioAIP are both open source tools for creating Archival Information Packages, or AIPS.
- AIPS are a critical component to any robust digital preservation plan.
- VideoAIP creates archival information packages from video files (like .MKV, .MOV, .DV, .MP4, etc.)
- Videoaip will incorporate into the AIP any .LOG or .FRAMEMD5 files that have the same name as your video input.
- AudioAIP creates AIPS from .WAV files.
- Is able to utilize webcam to take pictures of media carrier (or case), which often has relevant metadata you want to stay with your digital object. Includes these images in the AIP.
- **Linear Tape Open -** magnetic tape used for long term storage. Offline, very stable. Secure.
- Initial investment can be high, but once decks and equipment are purchased, cost is relatively stable.
- Decks and tapes are backwards compatible.
- LTOTools and LTOpers are two open source tools, used via command line, that can help you Format, Read, Write, and Verify the contents of your LTO tape.
- **MIPoPs basic file management -** Use basic [rsync](#) command to find new media in hot folders on our digitization computers, sync them to admin folder, then to [NAS](#).
- Create AIPS on admin computer, then write to LTO when we've accumulated enough data.
- Link together numerous microservices (many presented on [MIPoPS Microservices](#)) to build a system that is flexible and adaptable to our needs.

Resources for Digital Preservation Planning
- [National Digital Stewardship Alliance](#) (US

- ○ [NDSA Levels of Digital Preservation Matrix](#)
- [Northeast Document Conservation Center](#) (US)
  - ○ [Digital Preservation Assessment Training](#)
- [Digital POWRR](#) (Preserving digital Objects With Restricted Resources)
- [Library of Congress - Recommended Formats Statement](#) identifies hierarchies of the physical and technical characteristics of creative formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of creative content well into the future.
- [Digital Preservation Coalition](#) (UK)
  - ○ [Digital Preservation Coalition Rapid Assessment Model](#)
  - ○ [Digital Preservation Handbook](#)
- [PREMIS](#) - Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability.
- [MIPoPS' Resources Page](#)

Additional Tools
- VideoAIP - https://github.com/pugetsoundandvision/videotools
- AudioAIP - https://github.com/pugetsoundandvision/audiotools
- LTOpers - https://github.com/amiaopensource/ltopers
- LTOtools - https://github.com/pugetsoundandvision/lto_tools
- [Bagger](#) - open source application that packages data files according to the BagIt specification.
- [BagIt-Python](#) - open source Python library and command line utility for working with BagIt style packages.
- [BWF MetaEdit](#) - open source tools developed by FADGI which permits embedding, validating, and exporting of metadata in Broadcast WAVE Format (BWF) files.