



Embedded, Included or Needed -
Software and the Library of
Congress Tangible Media Project

Moryma Aydelott
Program Specialist, LS/TECH/O
Preserving.exe: Toward a National Strategy for Preserving Software
05/2013

[The Landscape]

- From the 2013 Survey – more than 250 TB in 16 Divisions
 - More than 100,000 CD items
 - More than 18,000 DVD items
 - More than 5,000 floppy discs (items such as 8", 5.25", 3.5", and 3" disks as well as Zip Drives and Bernoulli Disks.)
 - Almost 200 hard drives
 - Less than 100 flash drives

[The Project]

The Tangible Media project brings

- expertise,
- resources and
- training

to Library curatorial divisions that need assistance

- assessing,
- managing and
- preserving

their digital content stored on various kinds of physical tangible media.

[The Goal]

- A generic workflow for aggregation, bagging, transfer, inventory, access, verification and long term storage that can be used by multiple LS Divisions, various types of tangible media, and any digital formats.
- A centralized set of resources and help

[The Basics]

Who:

- LS Technology Policy led
- Working with seven LS curatorial and preservation Divisions

What:

- Over 1600 items (about 6.4 TB) moved off of tangible media onto Long Term Storage
- Most of the items are CD or DVDs, most of the data is from Hard Drives

Resources:

- About 2 years
- < 1 FTE (Project Manager, Technical and Curatorial Staff)
- 2 Ripstation jukebox machines
- Standard Library software

[In Relation to this Meeting]

We're doing good and relevant work but:

- we're not all software,
- we need software, and
- we don't have all of the solutions worked out

[The Work]

1. Bit Level Preservation and Triage
 - a) Transfer off of Tangible Media, saving both File and ISO versions.
 - b) Bagging using Bagger
 - c) Malware Scanning
 - d) Saving to Long Term Storage
2. Basic Access and Evaluation of Content
 - a) Right/Double-Click Access to Files
 - b) Validating, Opening, and otherwise Accessing Files As-Are
3. Additional Access, Migration/Reformatting/Emulation
 - a) Access through Catalog, Web Presentations
 - b) Migration/Reformatting of Files
 - c) Emulation of Software or Environment

[Software Preservation Issues]

- Processing Software On Items
- Using Software On Items
- Determining Software Needed
- Locating Software to Access Old or Obscure Formats

Processing Software On Items

Malware Scan Issues - LS/CS/HISP Reference Materials with IE

- Example at <https://cts.loc.gov/transfer/inventory/bag/50679>
- Approx 100 CDs, < 5 titles.
- Printed in the 1990's
- Collections of html and PDF docs, viewed in a web browser
- Processing workflows halted at Malware Scan stage, can't read included IE application files ; Malware scan of ISO image version had no problems.
- Solution: Since CDs work with web browsers installed on LC machines, removed the IE application file from the file version, retained it in the ISO, and left a note in the bag-info.txt file
- Preferred Solution: Confirmation if the issue is with malware in the software or a malware scan that needs tweaking, and an easy way to do it quickly.

2013-02-04 14:54:59	Service - Malware scan bag (Scan)	sun9servicequeue	Failed Cannot perform complete virus scan because not all files are readable: /sh_ingest/tangmed/hisp/00148847769 /data/00148847769/00148847769/IE4NT95/AMOV4IE.CAB /AMOVIE.OCX, /sh_ingest/tangmed/hisp/00148847769 /data/00148847769/00148847769/IE4NT95 /AMOV4IE.CAB/M
------------------------	---	------------------	--

[Details](#)

Using Software Already On Items

Working (but DOS) Applications - LS/CS/GM USGS DOQs and DRGs

- <https://cts.loc.gov/transfer/inventory/bag/2554>
- Approx 500 CDs
- Printed in the 1990's
- Collections of compressed GIS images and text metadata, additional functionality using DOS application
- Can use the images as-are, but need the DOS application to uncompress them.
- Solution: Immediate access to compressed images from File version of item, mount ISO version using mounting software to use DOS software.
- Preferred Solution: This actually works fine.



File Path	File Size (Bytes)	Options
DATA/		
DOCUMENT/		
DOQ.BAT	92	download identify
SOFTWARE/		
STARTUP.BPT	410	download identify
STARTUP.HLP	3,485	download identify

Determining Software Needed

Determining File Types - LS/CS/MUS and LS CS/GM Items

- With or without valid extensions, files occasionally can't be identified or validated using DROID or JHOVE.
- Can sometimes guess what kind of files they are based on collection and context (Box on right L and R files are likely WAV or other sound file. Similar issues with GIS data files in LS/CS/GM)
- Solution: Keep up with updates for file identification/validation tools, make best guesses, and try opening copies of files to see if respond as expected.
- Preferred Solution: List of recommended software available at LC to try for different types/ages of files.

File Path	File Size (Bytes)	Options
.DS_Store	6,148	download identify
66249-Again-Stereo.L	216,478,990	download identify
66249-Again-Stereo.R	216,478,990	download identify
66249-Again-Stereo.wav	432,956,582	download identify characterize

[Determining Software Needed]

Determining if the Software We Have is Good Enough or if the Provided Software is Required – LS/CS/GM NGA Items

- Tens of thousands of items (CDs and DVDs)
- Printed from 1990's – present.
- FOUO GIS data sets series, at the city, country and regional level.
- Items come with NGA-supported software for data transformations and conversions. Can do the same work with other tools we already have onsite, but it unclear what the downsides are.
- Solution: Keep up with updates for file identification/validation tools, make best guesses, and try opening copies of files to see if respond as expected.
- Preferred Solution: Have copies of multiple kinds of software available, especially those that are recommended for a particular format or source.

Determining Software Needed

Determining if the Software We Have is Good Enough or if Older Software is Required – LS/CS/MUS Items

- Tens of WordPerfect 4.2 documents
- Created in the late 1980's
- Lyrics and other documentation from LS/CS/MUS collections
- Collection did come with some WordPerfect application floppies that require a different operating system than what we have on normal LC machines. But can open the files in current word processing software.
- Solution: Using a copy, open up files and see if we can read text in them.
- Preferred Solution: Have copies of multiple kinds of software available, especially those that were used to create a particular item.

```
PresumptiveFormats:
  PresumptiveFormat (FormatIdentification):
    NativeIdentifier (I8R):
      Value: x-fmt/393
      Namespace: PUID
    JH0VE2Identifier (I8R):
      Value: http://jhove2.org/terms/format/wordperfect
      Namespace: JH0VE2
    IdentificationProduct (I8R):
      Value: http://jhove2.org/terms/reportable/org/jhove2/module/identify/DR0IDIdentifier
      Namespace: JH0VE2
    Confidence: PositiveSpecific
```

[Locating and Running Software to Access Old or Obscure Formats]

- Honestly, with the focus on bit-level preservation and mostly post-1990's materials haven't been forced to do a lot of this yet.
- But interested in doing more of this now, and expect to have to do a lot in the future.
- Issue is as much determining the software we need and having an environment that we can run in it in, as it is locating working installation disks.

[The End-Up]

- Project work is still in early phases
- For most projects, software is an external tool rather than what's being preserved
- Dealing with software and software preservation issues on case-by-case basis with the best tools, guidelines we have at the time
- Insights, lessons learned from elsewhere are welcome

[Questions?]

Contact:

Moryma Aydelott

Tangible Media Project Manager

mayd@loc.gov

202-707-1605

Online Resources (LC Access Only):

- Project Site: <http://www.loc.gov/staff/ls/techpol/reports-tangible.html>
- Complete List of and Links to Workflows:
<http://www.loc.gov/staff/ls/techpol/reports-tangible-workflow.html>