

Implementing Digital Preservation at the University of Houston

Bethany Scott, Coordinator of Digital Projects
UH Libraries, Special Collections

Presentation Outline

- History of digital preservation at UH
- Status of Archivematica implementation
- Workflows for accessioning and preserving born digital archives
- Providing access to digital objects
- Next steps

Digital Preservation at UH

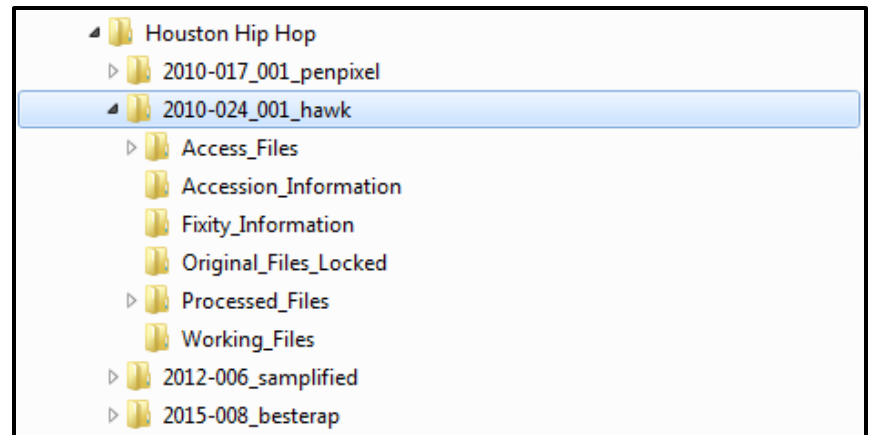
- Creation of Digital Preservation Task Force (DPTF)
- Conducting research
- Writing Digital Preservation Policy
- Testing and selecting Archivematica
- Restructuring the team

Archivematica Implementation

- Conducting inventories of digital assets to be preserved
 - Digitized vs. Born Digital
- Installation and configuration with Artefactual
- Set-up of Duracloud backups through TDL
- Developing AM workflows and doing test transfers
- Creation of Access & Preservation WG

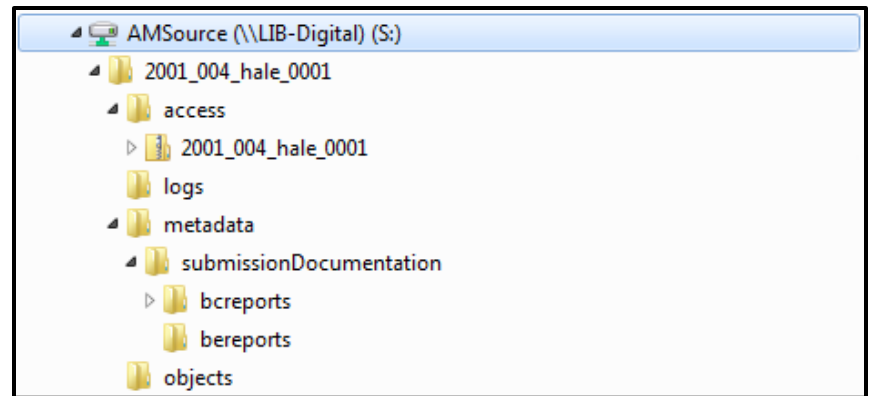
Previous Born Digital Workflow

- Acquire digital media objects
- Create accession folder on shared drive
- Copy files from removable media
- Place files and metadata/logs into accession folder
- Record admin info on transfers in digital media accession log
- Periodically back up copies of original files to the dark drive



Archivematica/BitCurator Workflow

- Add collection/item info to digital media accessioning log
- Create disk image using FTK Imager
- Run bulk_extractor and fiwalk reports in BitCurator
- Create descriptive metadata
- Structure package
- Transfer and ingest through AM
- Create digital object in Aspack and add package ARK



✓ 2001_004_hale_0001	bf68cde8-5b72-4e63-9a7f-9244581053df	2016-05-10 12:14	 
▶ Micro-service: Create SIP from Transfer			
▶ Micro-service: Complete transfer			
▼ Micro-service: Examine contents			
Job: Check for specialized processing	Completed successfully		
Job: Examine contents	Completed successfully		
Job: Examine contents?	Completed successfully		
Job: Move to examine contents	Completed successfully		
▶ Micro-service: Validation			
▼ Micro-service: Characterize and extract metadata			
Job: Load labels from metadata/file_labels.csv	Completed successfully		
Job: Characterize and extract metadata	Completed successfully		
▼ Micro-service: Update METS.xml document			
Job: Add processed structMap to METS.xml document	Completed successfully		
▼ Micro-service: Extract packages			
Job: Extract packages?	Completed successfully		
Job: Determine if transfer contains packages	Completed successfully		
Job: Move to extract packages	Completed successfully		
▶ Micro-service: Identify file format			
▶ Micro-service: Clean up names			
▼ Micro-service: Generate transfer structure report			
Job: Generate transfer structure report	Completed successfully		
Job: Move to processing directory	Completed successfully		
Job: Move to generate transfer tree	Completed successfully		
▶ Micro-service: Scan for viruses			
▶ Micro-service: Quarantine			
▶ Micro-service: Generate METS.xml document			
▶ Micro-service: Verify transfer checksums			
▶ Micro-service: Reformat metadata files			
▶ Micro-service: Assign file UUIDs and checksums			
▶ Micro-service: Include default Transfer processingMCP.xml			
▶ Micro-service: Rename with transfer UUID			
▶ Micro-service: Verify transfer compliance			
▶ Micro-service: Approve transfer			
✓ texas_city_disaster_objectsdironlymeta	59dcc85f-df6c-4f37-a2dd-312c7fa40b6f	2016-04-12 14:15	 
▶ Micro-service: Create SIP from Transfer			
▶ Micro-service: Complete transfer			
▶ Micro-service: Examine contents			

Processing configuration

Processing configuration

- General
- Failures
- Transfer source locations
- AIP storage locations
- Processing storage usage
- AtoM DIP upload
- Archivists Toolkit DIP upload
- PREMIS agent
- REST API
- Version

<input checked="" type="checkbox"/> Send transfer to quarantine	No ▾
<input type="checkbox"/> Approve normalization	Yes ▾
<input type="checkbox"/> Store AIP	Yes ▾
<input checked="" type="checkbox"/> Transcribe files (OCR)	No ▾
<input checked="" type="checkbox"/> Generate transfer structure report	No ▾
<input type="checkbox"/> Remove from quarantine after	<input type="text"/> days
<input type="checkbox"/> Create SIP(s)	--Actions-- ▾
<input type="checkbox"/> Extract packages	--Actions-- ▾
<input type="checkbox"/> Normalize	--Actions-- ▾
<input checked="" type="checkbox"/> Reminder: add metadata if desired	Continue ▾
<input checked="" type="checkbox"/> Examine contents	Examine contents ▾
<input type="checkbox"/> Select file format identification command (Transfer)	--Actions-- ▾
<input checked="" type="checkbox"/> Select file format identification command (Ingest)	Use existing data ▾
<input type="checkbox"/> Select file format identification command (Submission documentation & metadata)	--Actions-- ▾
<input type="checkbox"/> Delete packages after extraction	--Actions-- ▾
<input checked="" type="checkbox"/> Select compression algorithm	7z using bzip2 ▾
<input checked="" type="checkbox"/> Select compression level	5 - normal compression mode ▾

Providing Access to Digital Objects

- Digitized collections are accessed through UHDL
- Born digital materials may be accessed on reading room computer
- Creation of access files and transfer of files to shared drive
- Access notes for digital files are added to finding aids

Next Steps

- Continue refining born digital workflows
- Scale up transfers with move to production instance
- Develop utilities to prep packages for transfer
- Assess and update digital preservation policy

Thank you!

Contact me:

Bethany Scott

bscott3@uh.edu

713-743-0530