RDPP Meeting Notes 10/28/13

Digital Preservation @ Henry Ford by Brian Wilson
(brianw@thehenryford.org @brilwil)

Background on the Henry Ford
- Dedicated in 1929
- Edison Institute (current legal name)
- Five units
  - Archives & Library
    - Benson Ford Research Center
      - Public name for library & archives
  - Collections
  - Kennedy Limo
  - Wright Cycle Shop
  - Ford Motor Co. Records
    - 1903-1950s
  - Repository for Institutional records
    - Paper up to the 90s
    - Now electronic
    - Library

Digital Assets
- History
  - Film to disc in early to mid 1990s
  - Scanning regularly since 2002
- Sources
  - Scanning
  - Photography
  - Audio & film
  - Born digital
  - Donations
  - Institutional records
  - They get a lot of sales brochures
    - Multiple formats in each
- Formats
  - A little of everything
  - Using tiff as still image preservation format

- Amount
  - ~14TB in spinning storage
    - o adding 1.5TB each year or so
    - o mostly images, not text

Digital Preservation History
- Prior to 2010
  - burning backup copies to CDs
  - Portable hard drives
  - little or no selection
- Starting in 2010
  - Project launched to digitize "everything"
  - Focus on creation and presentation
  - Storage and preservation not considered
    - o trying to get more on board for long term preservation/storage
- Late 2010
  - inventory of assets
  - Using a networked RAID drive array
- Started with Policy
  - bit-level and format migration
  - language is starting to show up in grant applications
  - not formally approved yet
- Figuring out workflows
  - Created a pipeline based on OAIS

Real World
- Efforts have been crisis driven
  - Failed storage device
  - Depletion of network space
  - Blew out RAID drive holding 10 years of storage
  - recovered from backup CDs
- Procedures are partially implemented
  - ingest fixity check not in place
- Poor forecasting of space needs
  - ran out of network storage in Feb
    - o was only able to give IT a week to get more space
  - must be able to communicate
- Collection management system issues
    - o backup and preservation of db

- o does not control master files
- o no provision for business data
- o format control over attached files
- Limited visibility
  - no assigned staff role
    - o no agency
    - o no "cred"
  - lack of position in organizational structure
  - storage is not an asset
    - o would like storage visible storage for digital
      - ▪ not sure how that would work yet
    - o museum exhibit "what I can see" focus
    - o use of the term "paper" to describe archives
- Limited Resources
  - Staff, in terms of people, time, tech "comfort"
  - storage
- Limited Metrics and Reporting
  - focus on making stuff of a single genre
  - target for production, not preservation
    - o need metrics that expose the actual amount of work done
- Successes
  - Small successes
    - o daily imaging output store to network environment
    - o Kodak photo CD content migration
    - o File migration from portable hard drive to network
  - Improving Visibility
    - o IT institution storage review project
    - o grant applications include preservation needs
    - o continually raise topic in discussions and meetings
      - ▪ business storage is not the same as archival storage
  - o Resources
    - o training staff to free up time for dig pres activities
  - o Explore Cooperation/Sharing
    - o Local practitioner meetings
    - o Regional digital repository
- Summary
  - Better than 2010 but not good enough

- Partially identified materials and problem areas
- Developed drafts for policy and high-level processes
- Partially implemented procedures

Questions

What's your network like?

- large stack of virtualized servers
  - windows-based
- will run out of space in 4 weeks but getting another 1.5TB
- majority is backed up incrementally
- tapes stored elsewhere in Dearborn
- primary copy is on disc and accessible across network
- Windows permissions control
  - only 3-4 people can write, no one can delete without IT assisting
  - files backed up in business style IT backup, not archival level
- Storage cooperative could offer alternate ways of storing data in archival way, more geographic dispersement
- Cost is still an issue
  - Not necessarily the hardware, although it's not coming down
  - Management of files

What's the main challenge?

- Getting upper levels on board with their goals

Backlog in physical material?

- Not sure of extent of digital material that needs to be worked on

Still imaging

- managing that
- working on cutting down on metadata from 38 fields
- trying to cut down on time spent per record
- Flickr site
  - 45,000 images from one collection
  - people can still find what they need

Collection management system?

- EMU made by KE software
- Used for metadata
- one of the top systems for describing items in museums
- Subset is used like a large curated exhibit
- This is what is accessed by the public

Lots of micro-sites from past years

- Need to gather them together for use/content

Started a quarterly Web crawl

Interesting 3D items in collection

- Doughnut of Death

Results from possible workflows seminar survey
- All day workshop focused on digital pres/workflows
  - UM Library could host if there's interest
  - Created survey to capture ideas and interest
- Survey posted approximately 2 months ago
  - 23 responders- all enthusiastic
  - meeting happening in winter/spring 2014 (not 2013!)
    - few votes for early summer 2014
  - Suggested topics:
    - needs assessment
    - workflows for images
    - designing a workflow
    - how tools fit together
    - training
    - capture, metadata creation, and file management
    - migration
    - Accessioning and processing of digital content
    - Solutions for smaller institutions
    - Costs
    - Necessary infrastructure
  - Suggested tools for review
    - "all of them"
    - General tools for metadata production, checksums, PDF, video, processing
    - Repository software: Fedora and other duraspace tools, ArchiveSpace (a lot of interest), Islandoria (packaging of Fedora and Drupal)
    - Other tools: Oxygen (xml editing), Archivematica (a lot of interest)
  - Topics people offered to present:
    - needs assessment
    - locally developed workflows
    - capture, cataloging, storage, and delivery workflows
    - Getting started with institutional digital preservation (could be multiple sessions)
- Additional topics?

- - Records management
  - Web capture (maybe without using a crawler)
- Want to be careful about crossover with other DP orgs in the state
- Possibly have the symposium in May?
- Could do additional RDPP meetings around a specific topics not covered in the symposium
  - Ex: Web archiving
  - Could ask the Digital Curation group from the Bentley
- Symposium program committee
- UMSI/WSU students could help
- Lance wants volunteers to help him design content
  - Has some ideas
  - Wants to open it up to the community
  - Will put it on the Listserv

Topics for future RDPP meetings
- Metadata was the most popular suggestion on the survey
- How people are combining digital & paper collections?
- Advocating internally for financial support for preservation
- Regional partners for archiving and digital preservation
- Update on community storage project from Brian Wilson
- What NOT to save
- Fail Camp
  - People talking about their failures (i.e. data loss)
  - Could anonymize?

Next meeting
- January or February 2014

Tour of the Benson Ford Research Center facility