Practical options for incoming digital content

Alissa Matheny Helms & Jody DeRidder
62
Total survey responses.

27
Made it to the end.

100%
Useful information.

THANK YOU!
Type of role:
- Digital archivist
- Archivist
- Librarian/archivist
- Digital librarian
- Librarian
- Metadata librarian
- Manager
- None
Type of content collected

- Executable files
- GIS Data
- Email
- Databases
- Websites
- Moving images/videos
- Audio recordings
- Still images
- Digital text

The chart shows the percentage distribution of different types of content collected, with Digital text having the highest percentage at 100%.
Type of content collected

- Our 2016 Survey
- ARL 2012 Survey
- OCLC 2009 Survey
**File management**

**Preservation**
- 83% – preserve original files
- 17% – “sometimes or under certain circumstances”

**Normalization**
- 48% – normalize or migrate files
- 41% – “sometimes or under certain circumstances”
- 11% – do not normalize files
Technical metadata collected

- Structure of document
- Associated files
- Creating software (and version)
- File type version
- Original directory location
- Checksums
- File sizes
- File type
- File date(s)
Rights information collected

- Rights to make metadata
- Digitization permissions
- Rights to preserve & migrate
- Rights to make copies & derivatives
- Intellectual property rights
- Copyright
- Access & use restrictions
Rights information tracking

- 31% through a repository or finding aid
- 15% through an accession record
- 15% through a catalog record
- 46% through embedded metadata
  - DACS
  - METS
  - Dublin Core
Content organization & tracking

- Directory structure
- Label
- Collection & date
- Accession records
- CMS/Repository
- Spreadsheets
- Tools/Software
Target archival formats

- **Text documents**: 96%
  - PDF/A – 61%

- **Still images**: 83%
  - TIFF – 95%

- **Audio recordings**: 71%
  - WAV – 70%

- **Spreadsheets**: 71%
  - CSV – 70%

- **Moving images/video**: 58%
  - MPEG-4 – 57%

- **Databases**: 29%
  - CSV, TXT, original format – 29%
What works?
What doesn’t work?
Where in the world to start?
Digital content intake

Identification → Analysis → Selection → Processing
Identification

- examine what is there
- build an inventory
- identify duplicate files
- type of material
Identification – **useful tools**

- Karen's Directory Printer
- QuickView Plus
- DROID
- Spreadsheets
- In-house scripts
- TreeSize
- BitCurator
- FTK
Identification – **NOT useful tools**

**BitCurator**
- Better for disk image than individual files
- Too complicated; FTK more basic

**Data Accessioner**
- Did not fit workflow
- Replaced individual use tools
Analysis

- locate (and associate) all the versions of a particular file
- how many of each file type
- sets of files/file structures = single item
- system files, commonly used software, PII files
Analysis – useful tools

bulk_extractor
FTK
DROID
BitCurator
spreadsheets
Selection

- determine which of multiple versions of a file are of interest
- isolate files of particular types which the donor has specified
- sort through many types of files in many directories
- descriptive metadata generated or collected
Selection – **useful tools**

- Knowledge of collections
- PII Identification tools
- Retention schedules
- Collection policy
- FTK
- Manual review
Selection – **NOT useful tools**

BitCurator

- Not specific enough; disc level rather than file level
Processing

- accessible for current and future users
  - access points with current technology
  - preserve across generations of technology
    - content migration and data normalization
    - emulation implementation
Processing – useful tools

- Archivematica
- FTK
- Adobe Acrobat Pro
- Spreadsheets
- In-house scripts/software
- Handbrake
Processing – NOT useful tools

BitCurator
- Too complicated

MS Excel

Moodle and Omeka
- File format and size limitations
- Lack of public access in LMS
Recommendations from experienced digital content wranglers
Don't paint yourself into a corner – build an **infrastructure** that will grow and change with the types of content you receive.
Just start trying some of the **tools**... The more you work with them, the easier they will be. But make sure they work for your situation, not all tools work for all institutions.
Be guided by access in deciding how to arrange and process records.
Getting **involved** with the format specific organizations was very helpful early on... Digital librarians/curators/archivists need **our own association** to bring all the aspects together.
Discussion topics?

1. Technical Metadata
2. Rights Metadata
3. Identification Tools & Workflows
4. Selection Tools & Workflows
5. Specific tools (BitCurator, FTK, etc.)
6. What Archivists Need in the Field
7. What questions to ask donors
Google Doc

Help us make this information even more useful:

Thanks!

Any questions?

You can find us at:
- amhelms@ua.edu | @alinthearchives
- jlderidder@ua.edu | @jlderidder
Credits

Thanks to those of you who took time from your busy schedules to fill out our survey!

Special thanks to Amy Chen for assisting with the session!

And thanks to SlidesCarnival for the presentation template.

References
