Standards & Procedures

AMA Digitization Workshop

Elizabeth-Anne Johnson
University of Manitoba
29 February 2016

Digitization standards

- Having standards for preservation helps to ensure that digital files can be preserved and accessed over time
- Library and Archives Canada's format guidelines are here: https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx
- NARA (American National Archives and Records Administration) has a guide to preferred and acceptable file formats on their website: https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html
- Archivematica (digital archiving software) lists their format policies here: https://wiki.archivematica.org/Format_policies

File formats

- File formats are the type of file you're creating; different file formats have different attributes and are handled by computers differently
- Archival preservation formats are TIFF (for images and scans), PDF/A (the archival version of PDF, used for some scanned textual documents), and FLAC or WAV (for audio files)
- There is no real consensus on a best format for video files—uncompressed AVI and MOV files may be used, as well as MPEG-4 files.
- More information: http://www.digitalpreservation.gov/formats/
- https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html

Access copies

- Because preservation-quality files are so large, it is helpful to have smaller access copies of the files that are easier to share and open
- Access formats are JPEG (for images), PDF (for textual documents), MP3 (for audio), and MP4 (for video).

Textual material

- Always digitize to standards appropriate for the accurate preservation of the information on the printed page
- Scan most material at 600 dpi and in 48-bit colour if available
  - These specifications ensure that the digital file will contain more information than lower-quality scans will
  - Even if a document is in black and white or greyscale, scanning it in colour means that any gradations in colour or quality will be preserved in the digital file
- Preferred file formats: TIFF or PDF/A for preservation and JPEG or PDF for access

Photographs & other images

- Maintain the original dimensions of the image or object—for photographs this may include a border around the actual photograph
- Scan most objects at 600 dpi and in 48-bit colour if available
    - The longest side of the image must be at least 3000 pixels across (i.e. smaller items like slides must be scanned at a higher resolution)
- Preferred file formats: TIFF for preservation and JPEG for access

Audio files

- Digitize to standards appropriate for the accurate preservation of the original audio
- Transfer digital audio at a minimum of 16 bits per sample, but ideally at 24 bits per sample
- Transfer digital audio at a minimum sample rate of at least 44.1 KHz, but ideally at 96 KHz
- Preferred file formats: FLAC for preservation and MP3 for access

Video files

- Digitize to standards appropriate for accurate preservation of the original audio and video
- 8-bit video is acceptable but 10-bit is preferred
- Formats: vary from institution to institution
  - Preservation: Uncompressed AVI, MOV, or MPEG-4
  - Access: MPEG-4

Scanning procedures

- Make sure that all staff/volunteers who will be doing scanning work know the correct procedures for your institution
- Keep glass on scanner clean using microfibre cloth or products recommended in the scanner's manual
- Keep USB and power cords untangled and unkinked
- Update scanner drivers as needed
- Have a staff member perform quality control on scans—make sure the images meet archival standards and are useable

Scanning different media

- Paper media can simply be scanned by placing the document face-down on the glass of the scanner within the scanning area
- Slides and negatives need to be scanned with light coming from both sides of the item; most scanners will have a cover on the inside of the lid that you can remove to reveal another light
- Most scanners will come with trays which you can put slides or negatives into; the scanning software will then recognize each image as a separate one
- Make sure the proper settings for negative/positive images are selected

Naming conventions

- Take care to choose a naming convention that works for your institution
- Important things to keep in mind are the ease of sorting files within a folder, of identifying images as part of a collection without having to open them, making it simple for researchers to accurately cite you as a source
- U of M filenames reflect the collection number, accession number, box, folder, item, and page of the digitized item, e.g. UM_pc142_A97-45_001_0001_034_0004

Maintaining digital files

- Track when digital files were created—either in the filename or in the item's metadata
- Keep multiple copies of all digital media
- Keep storage media (hard drives/optical media) in stable and safe environments
- Optical media (CDs and DVDs) may degrade within 5 years
- Be aware of changing file formats and migrate digital material forward into the current preservation/access standards