Managing Digitization Projects

10 March 2010, INF2308H

Andrea Kosavic

Digital Initiatives Librarian, York University
Introduction to Digitization

- Why digitize?
- Digitization challenges
- Managing a digitization project
- Digitization:
  - Text and images, audio, moving images
  - Where to go for help
- Platforms and collaborative opportunities
- Metadata
- The Inuit through Moravian Eyes
Why Digitize?

- Obsolescence of source devices (for audio and moving images)
- Content unlocked from a fragile storage and delivery format
  - More convenient to deliver
  - More easily accessible to users
  - Do not depend on source device for access
- Media has a limited life span
- Digitization limits the use and handling of originals
Why Digitize?

- Digitized items more easy to handle and manipulate
- Digital content can be copied without loss
  - Analog formats degrade with each use and lose quality when copied
- Can be delivered to a far reaching audience over internet
- Can add metadata (enhances searching)
Digitization challenges

- Multiple formats to choose from
  - Formats constantly evolving
- Can’t match quality to that of the source
- Preservation challenges
  - Analog version is still considered the preservation master copy
- Expensive
  - Equipment, storage, metadata, staff time, long term preservation
Digitization challenges

- **Storage**
  - CD quality audio = 520 MB per hour
  - DVD-quality video = 13 GB per hour
  - Broadcast quality video = 75 GB per hour (ITU-R BT.601)

- **Technical limitations**
  - Compression algorithms still evolving
  - High bandwidth required for transfer
    - At preservation standards, it takes 5x the duration of an audio file to transfer over T1 network
Managing a Digitization Project
Project Planning

- What are your aims and needs?
- What do your users need? Try to integrate their feedback at all stages.
- What does administration want? Does this mesh with their aims?
- Distinguish between these needs, prioritize them, and create a plan.
Know your collection

- What do you want to scan?
- Will you be selecting specific items, if so, what’s your criteria?
  - Condition of originals
  - Copyright status
  - Items in high demand
  - Subject knowledge of selector
- Need estimated numbers
Minimize duplication of effort

- Check to see whether the items you wish to digitize have already been digitized

- Places to check:
  - WorldCat
    - Special instructions to search the Registry of Digital Masters here: http://www.oclc.org/services/collection/default.htm
  - Internet Archive
  - Early Canadiana Online
Digitization is a team effort

- Ensure you have the required support (departments, administration) and resources
- Collection knowledge is just as important as technical knowledge
- Plan for staff recruitment, training and attrition
- Keep channels of communication open
  - Problem solving has to be timely
Digital capture

- Establish file naming conventions and directory structure
- Conduct a small pilot study to test your workflow and settings
- Identify special handling requirements for materials and put in place appropriate guidelines and training
- Document the workflow and encourage team feedback
- Watch out for ‘noise’
Metadata

- Metadata is time consuming
  - Determine quality benchmarks
  - Can be an iterative process
- Determine how you want your collection to be searched and displayed
- Adopt controlled vocabularies
- When adapting formal metadata standards, ensure that you are not sacrificing interoperability
Outsourcing

- Get a trusted referral if possible
- You need to know technical details and standards to ensure that you get what you need
- Don’t forget about metadata
- Clarify what the price covers and how it breaks down
- Your agreement should include timelines and penalty clauses, quality assurance standards and procedures, and reporting requirements
Quality Assurance (QA)

- Establish clear criteria and well-documented quality assurance procedures
- Be realistic
- Allow adequate time to undertake QA and any corrective work
- Enable your users to alert you to any errors and provide you with evaluative feedback
- Evaluate as you go along and integrate what you learn into your project
Collection delivery

- Think about your interface at the beginning to ensure adequate digital and metadata capture
- Note that your content/metadata will need to outlive any current management system
- Involve your users in interface design and testing
- Address issues of usability and accessibility
- Support standards for dissemination and interoperability
Preservation and Maintenance

- Talk to your IT support people about file storage and software upgrades
- Put in place a strategy for preservation, identifying how often your collection should be backed-up, checked, and migrated
- Fully document the project to ensure understanding of all aspects: digitization and metadata standards, copyright status, system architecture
Digitization of Text and Images

- Choosing a scanner
- The Digitization Process
- Common Image Formats
Scanners are format specific

- Take inventory of what needs to be scanned, and the composition of your collections
- Choose the scanner that best suits the largest volume of your materials
  - Maps
  - Plans
  - Manuscripts
  - Plain Text
  - Drawings
  - Paintings
  - Photographs
  - Negatives
  - Microfilm
  - Transparencies
  - Slides
  - Charts & graphs
Flatbed Scanner

- Good for items such as smaller maps, drawings, plain text, etc.
- Auto sheet feeder attachments allow for fast digitization of single sheets
- Scans a variety of resolutions
  - 1 bit (black and white)
  - 8 and/or 16 bit (grayscale)
  - 24 and/or 48 bit (colour)

Image: http://content.answers.com/main/content/img/CDE/CREOSCN.JPG
Flatbed Scanner Tips

- Scan plain black and white text at 1 bit, this avoids grey background
- Scan black and white drawings with shading at 8 bit, or 1 bit with half-toning
- Scanning colour images with text is difficult, if scanning at 24/48 bit, text quality will suffer, will have to play with settings or scan separately
Digital Cameras

- Ideal for maps, plans, manuscripts, drawings, paintings
- Labour intensive for individual scans, high quality
- Book cradle keeps pages flat without damaging the book

Images:
Specialized Scanner Types

- **Microfilm scanner**
  - Specialized for microfilm

- **Slide/Negative scanner**
  - Higher resolution capture
  - Come with specialized cartridges to hold different sizes of film

- **Photo scanner**
  - Higher resolution capture

Images:
- [http://www.bearclover.net/epson-scanner/silverfast.html](http://www.bearclover.net/epson-scanner/silverfast.html)
Automated Book Scanner

- Hundreds of pages per hour
- Must be supervised
- Used for large book scanning projects
- Not suitable for rare or fragile materials
- Does not create preservation grade images

Targets for scanning

- Many different sizes and types available
- Scanned with image
- Help to calibrate colour balance for scan
- Use scanning software to create white and black calibration with target for each scan
- Saved with archival digital master
- Derivatives are usually made with the target cropped out

Image: http://www-rcf.usc.edu/~gainer/impa/imaging/kodak_q_60_example.jpg
Targets for scanning

Patches in columns 1-3, 5-7, and 9-11 have colors that follow published IT8 REFLECTIVE aim points. All other patches are approximations. See companion text for details; use at your own risk. PFink 11/95

http://www.imagequality.com/dtp/images/elec.it8.refl.jpg
Image Processing

- De-skew
- De-speckle
- Reduce background
- Rotation
- Register

Warning

- Only de-speckle and reduce background on images if absolutely necessary
- Processing often results in image quality loss
OCR Notes and Recommendations

- Do not compress TIFFs, this can lead to incompatibilities
- Adjust brightness and contrast so that text is as dark as possible and background is as light as possible (using a copy of original)
- Skew in text will compromise OCR
- OCR tends to be less reliable with headings
- OCR tends to not be corrected
OCR Notes and Recommendations

- Require special ‘zoning’ algorithms for text in column format, i.e. magazines
- Some OCR programs have a maximum pixel width of file
- OCR will not recognize handwritten script
- Special OCR programs are available for Gothic script, i.e. ABBYY FineReader7
## Sample Imaging Requirements

Table 1: Digital Master Image Files—Recommended Imaging Requirements

<table>
<thead>
<tr>
<th>Document Type</th>
<th>Resolution</th>
<th>Bit Depth</th>
<th>Enhancements Allowed</th>
<th>File Format</th>
<th>Compression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printed Text†</td>
<td>600 dpi</td>
<td>bitonal</td>
<td>Sharpening, descreening, cropping, deskewing, and despeckling</td>
<td>TIFF 5 &amp; 6</td>
<td>Lossless compression (e.g., ITU-G4)</td>
</tr>
<tr>
<td>Rare/damaged printed text</td>
<td>400 dpi</td>
<td>8-gray or 24-color</td>
<td>Contrast stretching&lt;br&gt;Minimal adjustments for tone and color</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., LZW)</td>
</tr>
<tr>
<td>Book Illustrations</td>
<td>400 dpi / 600 dpi with enhancement</td>
<td>8-gray or 24-color</td>
<td>Contrast stretching&lt;br&gt;Minimal adjustments for tone and color&lt;br&gt;Descreen/rescreen, sharpen</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., ITU-G4, LZW)</td>
</tr>
<tr>
<td>Manuscripts</td>
<td>300-500 dpi</td>
<td>8-gray or 24-color, if color is present in the original</td>
<td>Contrast stretching&lt;br&gt;Minimal adjustments for tone and color</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., LZW)</td>
</tr>
<tr>
<td>Maps &amp; other oversized items</td>
<td>300-400 dpi</td>
<td>8-gray or 24-color</td>
<td>Contrast stretching&lt;br&gt;Minimal adjustments for tone and color</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., LZW)</td>
</tr>
<tr>
<td>Graphic Art</td>
<td>400-600 dpi</td>
<td>8-bit/ channel internal reduction</td>
<td>Contrast stretching&lt;br&gt;Minimal adjustments for tone and color</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., LZW)</td>
</tr>
<tr>
<td>Photographic Prints</td>
<td>400 dpi</td>
<td>8-bit/ channel internal reduction</td>
<td>Contrast stretching&lt;br&gt;Minimal adjustments for tone and color</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., LZW)</td>
</tr>
</tbody>
</table>

http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf
Sample Imaging Requirements cont’d

<table>
<thead>
<tr>
<th>Works of art on paper</th>
<th>400 dpi</th>
<th>8-bit/ channel internal reduction</th>
<th>Contrast stretching</th>
<th>Minimal adjustments for tone and color</th>
<th>TIFF 5 &amp; 6</th>
<th>Uncompressed or lossless compression (e.g., LZW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparencies</td>
<td>4000-5000 on long end or 400 dpi on output &gt; 8” x 10”</td>
<td>8-bit/ channel internal reduction</td>
<td>Contrast stretching</td>
<td>Minimal adjustments for tone and color</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression; (e.g., LZW)</td>
</tr>
<tr>
<td>Microfilm</td>
<td>600 dpi blown back to original size</td>
<td>Bitonal 8-bit gray</td>
<td>Sharpening, descreening; cropping, deskewing, and despeckling</td>
<td>TIFF 5 &amp; 6</td>
<td>Uncompressed or lossless compression (e.g., ITU-G4, LZW)</td>
<td></td>
</tr>
</tbody>
</table>

2 Although 600 dpi 1-bit is a de facto standard for printed text, a comparable or richer text file may be produced in grayscale at 400 dpi.

3 Random or irregular halftones and those produced in color may be imaged at lower resolution, e.g., 300 because there is a lower incidence of moiré. It is recommended that high quality book illustrations, such as aquatints, collotypes, and engravings, especially those produced as separate plates, be retained for their artifactual value.

http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf
Scanning Formats

Digital Master
- TIFF format
- Resolution of 600 dpi/ ppi widely adopted for most materials
- Lower resolutions may be used to keep file sizes down for materials such as maps
- Bit depth depends on type of material

Web Delivery
- JPEG, JPEG 2000 (scalable)
- GIF only captures 256 colours
Digitization of Audio

- The Digitization Process
- Audio Formats

http://www.webbasedprogramming.com/Tricks-of-the-Java-Programming-Gurus/f4-1.gif
Audio Digitization Setup

- Playback device
  - With audio out (ideal)
  - OR professional microphone (only as last resort)
- Analog to digital converter
  - Aka: capture device
- Mixer (recommended)
- Computer with audio digitization software
- Headphones
Analogue to digital converters

Internal computer sound card

- prone to electrostatic interference from computer circuitry
- Often built from inferior quality components

http://www.techexcess.net/images/products/other/sb0200_medium.jpg
Analogue to digital converters

External capture device

- Provides superior results to sound cards

http://www.synthman.com/midiman/117212.html
Mixers (soundboards)

- Available as hardware and/or software
- Enables you to optimize the quality of your audio capture
Setup without external mixer

- Playback device with audio out connected to ADC which digitizes the audio.
- The digitized audio is then recorded by the computer.
- A software mixer installed on the computer can be used to mix the audio.
Setup with external mixer

- The playback device with an audio out is connected to a mixer where sound is optimized.
- The mixer is connected to the ADC which digitizes the audio.
- The ADC is connected to a computer which records the digitized sound.
Setup without audio output

- For playback devices that do not have an audio output, an external microphone can record the sound which can then be digitized...
Sampling Rate & Precision

- **sampling rate** = how many samples of sound are taken per second
  - at 96 kHz, sound is sampled 96,000 times per second

- **precision** is calculated in bits
  - the more bits a sample contains, the better the sound quality
  - 24 bit sample: 010011111100111001001101
Audio Preservation Standards

Sampling rate: 96 kHz
Precision: 24 bit
Encoding: Linear Pulse Code Modulation (LPCM) (not compressed)
Wrapper: Broadcast Wave Format (.bwf) or AIFF
Stereo encoding preferred over surround sound
(unless essential to creator's intent)

Notes:

- IASA (International Association of Sound and Audiovisual Archives) minimum recommendation for analogue originals is 48 kHz/24 bit
- DVD quality is 96 kHz/24 bit
- CD quality is 44.1 kHz/16 bit

http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
http://www.iasa-web.org/IASA_TC03/TC03_English.pdf
WAV vs BWF

- WAV files contain an info portion that is not governed by standards
- Broadcast Wave Format is a European standard created to append standardised metadata to the WAV audio file format
- BWF work on WAV players
Use and access copy

- Need proprietary software to play preservation master copies (96 kHz/24 Bit files)
  - Create CD with 44.1kHz/16 Bit file in .wav or .bwf format
- Web Accessible Copy
  - MP3
  - RealAudio, Quick Time (for streaming)
Use and Access Copy

 Original remains untouched
  • “Imperfections” may be significant to historians
 Copies may be enhanced by filtering and noise reduction techniques
  • Remove hiss, clicks and pops
  • Adjust calibration and EQ curves to approximate signal characteristics of original
### Table of standard audio formats

<table>
<thead>
<tr>
<th>Wrapper Formats</th>
<th>File Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Authoring Format (AAF)</td>
<td><strong>Compressed</strong></td>
</tr>
<tr>
<td>Advanced Systems Format (.asf)</td>
<td><strong>Uncompressed/Lossless Compression</strong></td>
</tr>
<tr>
<td>Audio Interchange File Format (.aif; .aiff) – preservation standard</td>
<td><strong>Advanced Audio Coding (.aac; .m4a)</strong></td>
</tr>
<tr>
<td>Audio/Video Interleaved (.avi)</td>
<td><strong>Compact Disc Audio (CDDA)</strong></td>
</tr>
<tr>
<td>Broadcast Wave Format (.bwf) – preservation standard</td>
<td><strong>Digital Audio Compression (AC-3; Dolby Digital)</strong></td>
</tr>
<tr>
<td>Jpeg 2000 (JP2)</td>
<td><strong>Linear Pulse Code Modulated Audio (LCPM) – preservation standard</strong></td>
</tr>
<tr>
<td>MPEG-4</td>
<td><strong>MPEG-1 Layer-3 (.mp3)</strong></td>
</tr>
<tr>
<td>MPEG-7</td>
<td><strong>Real Audio (.ra; .rm; .ram)</strong></td>
</tr>
<tr>
<td>MPEG-21</td>
<td><strong>Standard Musical Instrument Digital Interface (MIDI) File (.smf; .mid)</strong></td>
</tr>
<tr>
<td>Material Exchange Format (MXF)</td>
<td><strong>Windows Media Audio format (.wma)</strong></td>
</tr>
<tr>
<td>OGG format (.ogg)</td>
<td><strong>Wave (.wav)</strong></td>
</tr>
<tr>
<td>Quicktime (.mov, .moov, .qt)</td>
<td><strong>Extensible Media Format (.xmf)</strong></td>
</tr>
<tr>
<td>Real Media (.rm)</td>
<td></td>
</tr>
</tbody>
</table>
Digitizing Moving Images

- Challenges
- The Digitization Process
- Moving Image Standard Formats
Challenges

- Correctly identifying the material
- Understanding how the material was meant to be played back (e.g., frame rate)
- Finding a compatible play back device:
  - In good working order
  - Within budget
  - With service professionals available
  - With extra parts available
Digitizing Videotapes

- Videotapes are inserted into an analogue playback device (VCR)
- VCR is connected to an analogue to digital video converter
- The analogue to digital video converter is then connected to the computer which is loaded with special software that enables video capture

Image: http://www.grassvalley.com/assets/media/2288/twinpact100.png
Videotapes are played on an analogue playback device

The video signal is optimised using a time base and a colour corrector

The video signal is digitised using a desktop computer. The process is monitored using a cube monitor and headphones

The digital video files are archived onto hard drives and transcoded for distribution

Image 1.7 Typical analogue videotape to hard drive equipment configuration (click for larger version)
Automation options

- **SAMMA Solo or SAMMA Robot**
  - Vendor: Front Porch Digital
    - [http://www.fpdigital.com/Products/Migration/Default.aspx](http://www.fpdigital.com/Products/Migration/Default.aspx)
  - Enables thresholds to be sent so that videos can be transferred without supervision
    - Operator receives an email alert if an error has occurred
Digitizing Film

Multiplexer (Telecine)

- Image projected via lens and mirrors directly into camera
  - A camera captures another’s playback
- Quality suffers generational loss
- Generally used for film to videotape transfer or for television broadcasting of films
- Popular due to acceptable quality and affordability

Digital Film (Chain) Scanners

www.visinst.com/1635Photo2.gif
http://uk.gizmodo.com/flashscan8.jpg
Chain Film Scanner

- Digitize directly from 8, 16, or 35 mm
- Scans the film and digitizes at the scanner
- Passes the digital signal to the computer
- Digital conversion is done at the camera instead of computer
- Less opportunity for noise
- Expensive to acquire hardware
Reality check

- Local moving image digitization will likely be VHS centric
- Playback devices are difficult to obtain and maintain
  - Need specialists to properly run and maintain the devices
  - Need parts for upkeep
- For older formats, vendors will be a more viable option
Recommendations for digital master preservation

- Larger picture size preferred
- High definition content preferred (assuming picture size is equal or greater)
- Encodings that maintain frame integrity preferred over temporal compression
- Uncompressed or lossless compressed preferred over lossy compressed

http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
Recommendations for digital master preservation cont'd

- Higher bit rate (mb/s) preferred
- Extended dynamic range (brightness) preferred over “normal” dynamic range (for scanned motion picture film and Digital Cinema)
- Stereo and monoaural sound preferred over surround sound (surround sound only necessary if essential to creator’s intent)
Common moving image wrapper and file formats

<table>
<thead>
<tr>
<th>Wrapper Formats</th>
<th>File Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Authoring Format (AAF)</td>
<td>MPEG-1</td>
</tr>
<tr>
<td>Advanced Systems Format (.asf)</td>
<td>Uncompressed/Lossless compression</td>
</tr>
<tr>
<td>Audio Interchange File Format (.aif; .aiff)</td>
<td>Digital Cinema Initiative Distribution Master (DCDM)</td>
</tr>
<tr>
<td>Audio/Video Interleaved (.avi)</td>
<td>MPEG-2</td>
</tr>
<tr>
<td>MPEG-4</td>
<td>MPEG-4</td>
</tr>
<tr>
<td>MPEG-7</td>
<td>Uncompressed/Lossless compression</td>
</tr>
<tr>
<td>MPEG-21</td>
<td>Animation codec (Quicktime)</td>
</tr>
<tr>
<td>Material Exchange Format (MXF)</td>
<td>Real Video (.ram, .rm)</td>
</tr>
<tr>
<td>OGG format (.ogg)</td>
<td>SheerVideo</td>
</tr>
<tr>
<td>Quicktime (.mov, .moov, .qt)</td>
<td>Windows Media Video format (.wmv)</td>
</tr>
<tr>
<td>Real Media (.rm)</td>
<td>DivX (.divx)</td>
</tr>
<tr>
<td></td>
<td>Digital Video formats (DV, DVCAM, DVCPRO)</td>
</tr>
</tbody>
</table>

TABLE DATA FROM: http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
## Format Size Comparison

<table>
<thead>
<tr>
<th>Format</th>
<th>1 min video</th>
<th>1 hour video</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPEG1</td>
<td>10.4 MB</td>
<td>624 MB</td>
</tr>
<tr>
<td>WMV</td>
<td>12.4 MB</td>
<td>744 MB</td>
</tr>
<tr>
<td>AVI</td>
<td>214 MB</td>
<td>12 000 MB (12 GB)</td>
</tr>
</tbody>
</table>

Format recommendations for digital masters

Digital moving images (general case):
- .mjp or .jp2 inside a JPEG2000 wrapper

Digital video converted from analog tapes:
- MPEG-2 at a minimum data rate of 1 Mb/s
- MPEG-4 at a minimum rate of 0.5Mb/s

http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
Format recommendations for digital masters cont'd

High quality video (professional videotape):
- JPEG2000 uncompressed

Commercial movies:
- DCDM (Digital Cinema Initiative Distribution Master)

Digital broadcast television streams:
- Inconclusive, industry is in a state of flux

http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
Format recommendations for digital masters cont'd

- Note: Other preferred wrapper formats are AVI, QuickTime or WMV as long as audio and video bitstreams are uncompressed or use lossless compression.
Popular use and access formats

Streaming:
- Real Media Video
- Windows Media Video
- Quicktime
- MPEG-4 (multimedia)

Video CD:
- MPEG-1

DVD:
- MPEG-4

http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
Where to go for help with digitization questions...
Joint Information Standards Committee

Still images, moving images and sound advice

JISC Digital Media, formerly TASI, exists to help the UK’s FE and HE communities embrace and maximise the use of digital media and to achieve solutions that are innovative, practical and cost effective.

Free help and advice to the UK Further and Higher Education community

Helpdesk
JISC Digital Media Case Studies
http://www.jiscdigitalmedia.ac.uk/tags/category/case-studies/

JISC Digital Media mailing list
http://www.jiscdigitalmedia.ac.uk/mailing-list/

Association of Moving Image Archivists (AMIA) discussion list
http://www.amianet.org/participate/listserv.php

International Association of Sound and AudioVisual Archives mailing list
http://www.iasa-web.org/listserv.asp
Local Toronto Classes

- Henry’s School of Imaging
  - Scanning & Restoration
    (3 hours, $75)

- Vistek Seminars and Events
  - Scanning Negatives, Slides and Prints
    (3 hours, $75)
  - Basic Video Editing with Final Cut Pro
    (4 hours, $150)
PRONOM technical registry

- Holds information about file formats, and the software products which can process them
- Supports preservation efforts
- Search by file format, extension, vendor, software, lifecycle, migration pathway
- http://www.nationalarchives.gov.uk/aboutapps/PRONOM/tools.htm
Search: Simple Search

1. Search

Enter a simple search string and then click 'search'.

Search
Metadata

- Why create metadata?
- Types of metadata
- Systems & Schemas

```xml
<titleStmt>
<title>Survey for Rural Economies, 1998</title>
<IDNo>7856</IDNo>
</titleStmt>
<fundAg>Countryside Agency</fundAg>
<fundAg>Department of the Environment, Transport and the Regions</fundAg>
<copyright>Social Research Centre</copyright>
<grantNo>3289460</grantNo>
<distStmt>
<distbtr abbr="UKDA" affiliation="University of Essex, Wivenhoe Park, Colchester, Essex, England, CO4 3SQ">UK Data Archive</distbtr>
<depositr>National Centre</depositr>
<depDate date="2000-05-08"/>
<distDate date="2000-06-08"/>
</distStmt>
<keyword>ACCESS TO COUNTRYSIDE</keyword>
<keyword>AGRICULTURAL DEVELOPMENT</keyword>
<keyword>AGRICULTURAL PRODUCTION</keyword>
<keyword>AIR POLLUTION</keyword>
<keyword>COUNTRYSIDE</keyword>
<keyword>ENVIRONMENTAL CONSERVATION</keyword>
<nation>Great Britain national</nation>
<geogCover>GREAT BRITAIN</geogCover>
<geogUnit>(A)Wards; (B)Standard Regions; (C)Postcode Sectors; (D)Parliamentary Constituencies; (E)Local Authority Districts; (F)Counties; (G)Scottish Regional Councils</geogUnit>
```
Why do we need metadata?

- Digital identification
  - Used to differentiate one object from another
  - Used to identify sets of data
- Organizing e-resources
  - Organizing links to resources based on audience or topic
  - Building these pages dynamically from metadata stored in database
Why do we need metadata?

- **Resource discovery**
  - Allowing resources to be found by relevant criteria
  - Identifying resources
  - Bringing similar resources together
  - Distinguishing dissimilar resources
Why do we need metadata?

- Facilitating interoperability
  - Federated searching across collections
  - Allows for sharing and transfer of data
  - How?
    - Use defined metadata schemas
    - Share transfer protocols and crosswalks
    - Example: OAI protocol for Metadata harvesting
Why do we need metadata?

- Archiving and preservation
  - Digital information is fragile and can be corrupted or altered
  - It may become unusable as storage technologies change
  - Metadata is key to ensuring that resources will survive and continue to be accessible into the future:
    - track lineage/provenance
    - detail its physical characteristics and behavior in order to emulate it in future technologies
Types of Metadata

- Descriptive
  - Describes a resource for purposes such as discovery and identification
  - Can include elements such as title, abstract, author, and keywords

http://www.niso.org/standards/resources/UnderstandingMetadata.pdf
Types of Metadata

- **Structural**
  - Indicates how compound objects are put together
  - Example:
    - Show relationships between digital object and page number of book
    - The first scanned page of a book is rarely marked as page #1 of the book itself

http://www.niso.org/standards/resources/UnderstandingMetadata.pdf
Types of Metadata

- **Administrative (and Technical)**
  - Provides information to help manage a resource such as:
    - when and how it was created, file type and other technical information, and who can access it
  - **Subsets of administrative data:**
    - **Terms and Conditions**
      - deals with intellectual property rights
    - **Preservation Metadata**
      - contains information needed to archive and preserve a resource

Dublin Core

- Comes in a simple (15 elements) and a larger qualified set
- All elements are optional and repeatable
- Minimum standard for describing digital objects

### Simple Dublin Core Set:

<table>
<thead>
<tr>
<th>Title</th>
<th>Source</th>
<th>Contributor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Language</td>
<td>Date</td>
</tr>
<tr>
<td>Subject</td>
<td>Relation</td>
<td>Type</td>
</tr>
<tr>
<td>Description</td>
<td>Coverage</td>
<td>Format</td>
</tr>
<tr>
<td>Publisher</td>
<td>Rights</td>
<td>Identifier</td>
</tr>
</tbody>
</table>
Wrapper Formats

- Wrapper formats tie together many different types of metadata
- Ideal for preservation
- MPEG-21 and METS support moving images
- XML based
MPEG-21

- Specialized for preservation of moving images
- Allows detailed capture of intellectual rights info
- Very complex and hence only adopted by specialized archives

METS

- Metadata Exchange and Transmission Standard
- Created for describing complex digital library objects

Components of a METS File:
- METS Header
- Descriptive Metadata – MODS, MARC, MARCXML etc.
- Administrative Metadata – provenance and copyright
- Structural Map – hierarchy and links to digital objects
- Structural Links
- Behavior
MARC, MARCXML, MODS

- **MARC** (Machine Readable Cataloguing Record)
- Can easily transform:
  - MARC21 > MARCXML > MODS
- MODS is a subset of MARCXML elements
  - MODS is embedded in METS records for item level descriptive metadata
Sample Extension Schemas

- **Audio**
  - AudioMD, specific to audio e.g., channel or track specifications, sampling frequency.

- **Video**
  - VideoMD, specific to video files, e.g., bit rate, compression codec.
  - MIX, specific to images, e.g., bits per pixel, color space

- **Images**
  - ImageMD, specific to images e.g., type or condition
  - MIX, specific to images, e.g., bits per pixel, color space

- **Other**
  - RightsMD: Rights, restrictions, and/or other categorizing information that can be used to support rights-management and/or access-management systems.
  - ProvenanceMD: About the events/steps/processes that occurred in reformatting or migrating entities.

For more information: [http://www.loc.gov/rr/mopic/avprot/metsmenu2.html](http://www.loc.gov/rr/mopic/avprot/metsmenu2.html)
Recommended minimum metadata set for archiving moving image and sound resources

- Combines elements from Dublin Core, PREMIS, AudioMD, VideoMD, TVAnytime, MPEG-7

  - See pages 82 through 89 from: http://www.jisc.ac.uk/media/documents/programmes/preservation/moving_images_and_sound_archiving_study1.pdf
Dissemination

- Platforms
- Collaborative opportunities
Digital Collections Platforms

- Content DM (vendor)
- Greenstone, Kete, Omeka, Scribblio (open source)
- California Digital Library’s eXtensible Text Framework (XTF) (open source)
- Repository platforms: DSpace, Islandora (Fedora) (open source)
## All 270 items

<table>
<thead>
<tr>
<th>Title</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>.577 caliber percussion rifle-carbine, 1862</strong></td>
<td></td>
</tr>
<tr>
<td>Royal Small Arms Factory, Enfield, England</td>
<td></td>
</tr>
<tr>
<td>Tags: soldiers</td>
<td></td>
</tr>
<tr>
<td><strong>&quot;$100,000 reward!&quot; 1865</strong></td>
<td></td>
</tr>
<tr>
<td>George F. Nesbitt &amp; Co., New York</td>
<td></td>
</tr>
<tr>
<td>Tags: assassination, broadsides, conspirators</td>
<td></td>
</tr>
<tr>
<td><strong>&quot;10 Likely and Valuable Slaves at Auction,&quot; 1823</strong></td>
<td></td>
</tr>
<tr>
<td>Tags: broadsides, slavery</td>
<td></td>
</tr>
<tr>
<td><strong>&quot;100 Dollars Reward,&quot; July 6, 1857</strong></td>
<td></td>
</tr>
<tr>
<td>Tags: broadsides, slavery</td>
<td></td>
</tr>
</tbody>
</table>
Book Search

Repository Advanced Search
Total Hits = 2312, Number of Hits/page = 50
You may not have sufficient privileges to view any or all of the items found. The objects you have rights to view are:

1. llives:90627-z_004-img1p
   Score:(9.3404255)
   Photograph 1 - Back Matter 4 - It happened in Iona

2. llives:90627-p_116-img1i
   Score:(9.3404255)
   Illustration 1 - Page 116 - It happened in Iona

3. llives:90627-p_109-img1i
   Score:(9.3404255)
   Illustration 1 - Page 109 - It happened in Iona

4. llives:90627-p_105-img1i
   Score:(9.3404255)
   Illustration 1 - Page 105 - It happened in Iona

5. llives:90627-p_099-img1i
   Score:(9.3404255)
   Illustration 1 - Page 99 - It happened in Iona

Islandora http://islandora.ca/
<table>
<thead>
<tr>
<th>Image</th>
<th>Title</th>
<th>Date Published</th>
<th>Publisher</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ 1.</td>
<td>Carte qui contient une desoration des iles &amp; terres que les Anglios possedent dans l'Amerique Septentrionale</td>
<td>circa 1719</td>
<td>chez l'Honore B. Chataill Libraires</td>
<td>North Carolina Collection</td>
</tr>
<tr>
<td>☐ 2.</td>
<td>Carolina</td>
<td>1729</td>
<td></td>
<td>North Carolina Collection</td>
</tr>
<tr>
<td>☐ 3.</td>
<td>Carte de la Caroline meridionale et septentrionale et de la Virginie</td>
<td>circa 1770s</td>
<td></td>
<td>North Carolina Collection</td>
</tr>
<tr>
<td>☐ 5.</td>
<td>Carte de la Caroline et Georgia pour servir a l'Histoire generale des voyages</td>
<td>circa 1780</td>
<td></td>
<td>North Carolina Collection</td>
</tr>
<tr>
<td>☐ 6.</td>
<td>North Carolina</td>
<td>1795</td>
<td></td>
<td>North Carolina State Archives</td>
</tr>
</tbody>
</table>
Opportunities for collaboration...
We found 80 matching items.

**Media type(s):** Video

1. **Paul Laforet: An interview**
   - A video interview with Paul Laforet focusing on his experiences in the RCAF in World War II. A link to play the movie is on the right side of this page.

2. **Windsor Veterans' Project**

3. **Arthur V. Drake: An interview**
   - A video interview with Arthur V. Drake focusing on his experiences in the Canadian Army from 1943-1945

4. **Windsor Veterans' Project**

5. **Robert Murray: An interview**
   - A video interview with Robert Murray focusing on his experiences in the RAF in World War II (edited short version)

6. **Windsor Veterans' Project**

---

http://www.ourontario.ca/
A dictionary of the English language. By Samuel Johnson. A general history of all voyages and travels throughout the old and new world (1703). The history of the devil, as well ancient as modern (1727). Pastoral, epistles, odes, and other original poems, with translations from Pindar, Anacreon, and Sappho (1748). Esprit, maximes et principes de m. Jean-Jacques Rousseau, de Genève (1764). Twenty stories from Gismon (1896). A general history of all voyages and travels throughout the old and new world, from the first ages to this present time, illustrating both the ancient and modern geography, containing an accurate description of each country, its natural history, and...
MIC collects 558,489 catalog records from 15 participating institutions.
http://mic.loc.gov/
The Labrador Inuit Through Moravian Eyes

This site provides information on the 250-year relationship between Moravian missionaries and the Inuit of Labrador. This interaction led to the establishment of settlements for a formerly nomadic people, their conversion to Christianity and exposure to aspects of North American culture. The information has been gathered from a variety of sources that shed light upon this unique adventure. Read more >>

Black and white photograph of two Inuit children, circa 1927.
About the Project

- Canada Culture Online grant for 400,000+
- Collaboration between University of Toronto Libraries, Memorial University Libraries and the Bibliothèque de l'Université Laval
- Memorial University of Newfoundland provided source materials and description
- U of T responsible for digitization and interface
- Université Laval responsible for French translation
Types of Media

- Video
- Audio
- Photographs
- Drawings/Paintings
- Plans/Maps
- Manuscripts
- Published Texts
Additional Metadata for Browsing
Digitization Standards

- Photographs, Manuscripts, Plans/Maps, Drawings/Paintings
  - captured as 600 dpi 24 bit TIFFs

- Published Texts
  - 600 dpi 1 bit TIFFs.

- Delivered online as 3 sizes of JPEG
  - Thumbnail: 75 pixels across
  - Small: 500 pixels across
  - Large: 775 pixels across (to neatly fit inside borders of website)
Zooming Capabilities

- For Plans/Maps, we wanted to be able to show more detail.
- The Zoomify program was used.
- Zoomify takes an image and creates nested directories of tiles, only retrieving the tiles of interest.
- The result is slick and smooth zooming.
- This works like the zooming feature of JPEG 2000.
Scotiabank Information Commons

New Media Suites

- For use by UofT community
- Must complete free certification course
- Course teaches you how to use the equipment (about 2-3 hours)
- Have facilities for digitizing audio and video, scanners available as well
- Rent rooms for 3 hour time blocks

http://www.utoronto.ca/ic/newmedia/equipment.html
New Media Suites @ UofT

A/V Equipment in the Suites:
- Tascam 102 MK2 audio cassette recorder
- Pioneer DV-525 DVD player
- Panasonic 5710 SVHS video tape recorder
- JVC BR-DV3000 professional DV recorder

Software in the Suites:
- Avid Xpress Pro
- Adobe Photoshop
- Sorenson Squeeze
- Ulead DVD MovieFactory
Audio Items

- Digitized from audio cassettes at Scotiabank Information Commons in New Media Suites
- Digitized at 44.1 kHz, 16 Bit
- Used Avid Express Pro to capture and edit
  - Tape Player > ADC > Computer
- Pro Tools was used to boost gain where capture was not adequate
Basic Sound Recording Principles

- Must control input levels so that captured sound is not:
  - Too loud, otherwise clipping will occur
  - Too soft, otherwise you will have to process it to be louder

- We captured files too quietly, had to go back and boost levels
Example of a clipped wave
Example of a wave that needs boosting
Acceptable audio wave
Vendors

When money, time, equipment or expertise is short…

- Outsource to a trusted, recommended vendor
- This is often the most affordable and desirable option, especially for older formats
- Talk to your network of colleagues for recommendations
- Try to find a local vendor if possible
Moving Images

- Super 8 mm reels with sound
- Digitized to DVD (MPEG2) by trusted, local vendor
- Vendor recommended by Thomas Fisher Rare Book Library
- Digitization cost about $150 / reel
- Transferred from DVD into Avid environment for editing
The Real Work Begins

To ensure that capture was successful:

- Listened to each entire tape
- Watched each DVD
- Selected excepts from digitized audio and video for web
- Used Sorensen Squeeze to create derivative formats
- Digital masters saved in MPEG2 format
Web Delivery Formats

Video

- Quick Time and Windows Media
  - 256Kbps (56 Kbps was too blurry)
  - 512Kbps
  - 1Mbps

Audio

- Quick Time Audio and Windows Media Audio
  - 56Kbps
  - Broadband (128 Kbps)
Thanks for your time. Questions?