Introduction to Digitization: An Overview

June 3rd 2009, FIS 2308H

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, ie. MPEG7 allows enhanced searching

Digitization challenges

- Multiple formats to choose from

- Can't match quality to that of the source

- Preservation challenges

  - Analog version is still considered the preservation master copy

- Expensive

  - Digitization equipment, storage, staff time, long term preservation

Digitization challenges

- Storage…we're talking TBs!
  - CD quality audio is 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
    - For an audio file recorded at preservation standards, it takes 5x the duration of the file to transfer over T1 network

Managing a Digitization Project

Slides 9-17 adapted from: Learning Lessons from Other Digitisation Projects,
http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/learning-lessons-from-other-digitisation-projects/

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

- Need estimated numbers

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, it usually takes longer than digital capture

- Determine how you want your collection to be searched and displayed – this will inform what metadata you will need to capture

- 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, or migrated

- Fully document the project to ensure understanding of all aspects: digitization and metadata standards, copyright status, system architecture

Digitization of Text and Images

- Introduction to various materials
- The Digitization Process
- Common Image Formats

http://www.wpclipart.com/computer/hardware/scanner.png
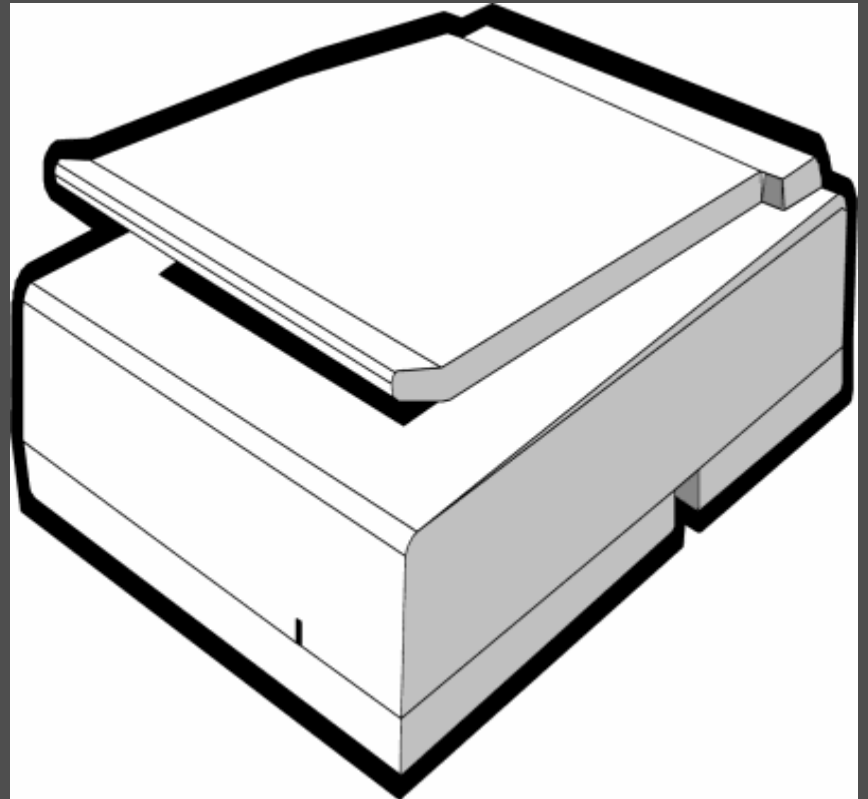
Multiple format types

- Maps
- Plans
- Manuscripts
- Plain Text
- Drawings
- Paintings

- Photographs
- Negatives
- Microfilm
- Transparencies
- Slides
- Charts & graphs

Flatbed Scanner

- Good for smaller plans / maps, photographs, plain text

- Auto Sheet Feeder attachments allow for fast digitization of single sheets

- Scans a variety of resolutions

- Scans at 1 bit (black and white), 8 bit (grayscale), and 24 or 48 bit (colour)

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 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 book

Specialized Scanner Types

- Microfilm scanner
    - Specialized for microfilm
- Slide/Negative scanner
    - Higher resolution capture
    - Come with specialized cartridges to hold different sizes of film
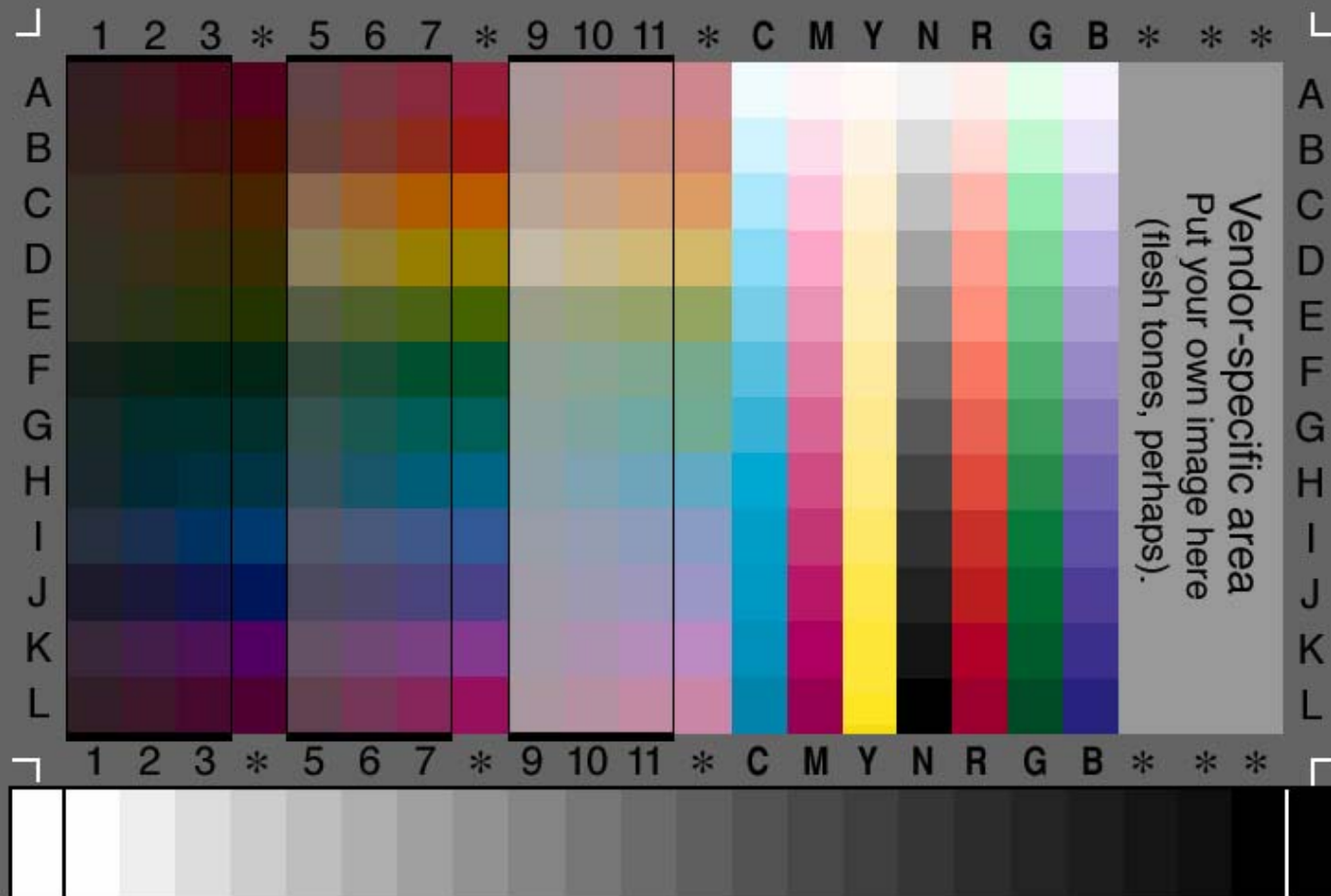- Photo scanner
    - Higher resolution capture
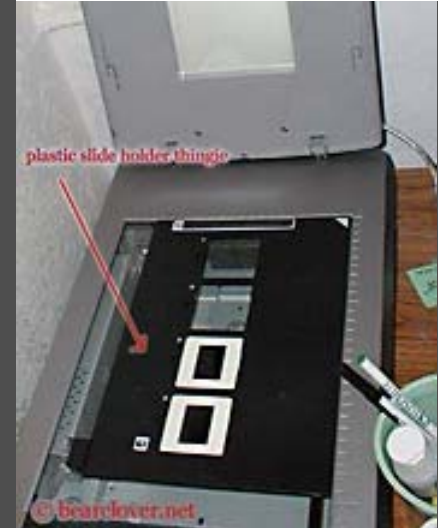
Automated Book Scanner

- 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

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

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

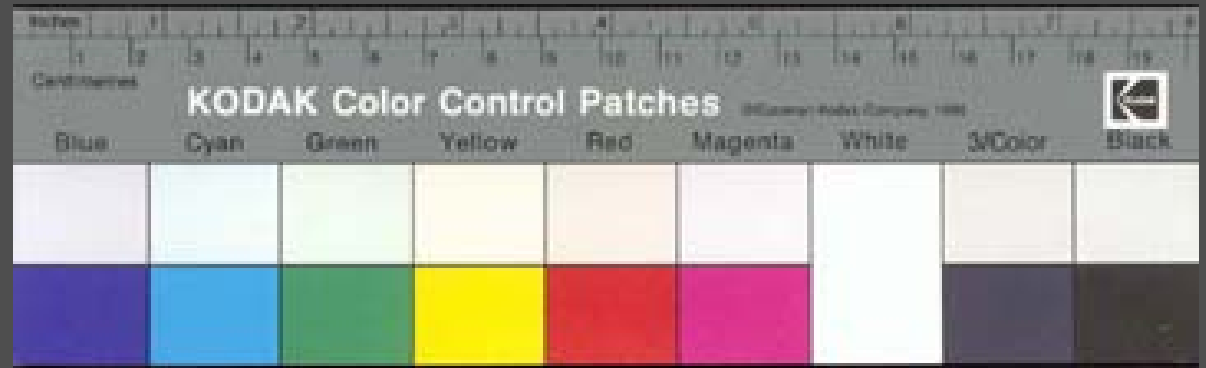- Derivatives are usually made with the target cropped out

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

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, incompatible with some OCR programs
- 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, ie. 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 ie. ABBYY FineReader7

Sample Imaging Requirements

Table 1: Digital Master Image Files— Recommended Imaging Requirements

Document Type	Resolution	Bit Depth	Enhancements Allowed	File Format	Compression
Printed Text[2]	600 dpi	bitonal	Sharpening, descreening, cropping, deskewing, and despeckling	TIFF 5 & 6	Lossless compression (e.g., ITU-G4)
Rare/damaged printed text	400 dpi	8-gray or 24-color	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Book Illustrations	400 dpi[5] 600 dpi with enhancement	8-gray or 24-color ----------- bitonal	Contrast stretching Minimal adjustments for tone and color ------- Descreen/rescreen, sharpen	TIFF 5 & 6	Uncompressed or lossless compression (e.g., ITU-G4, LZW)
Manuscripts	300-500 dpi	8-gray or 24-color, if color present in the original	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Maps & other oversized items	300-400 dpi	8-gray or 24-color	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Graphic Art	400-600 dpi	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Photographic Prints	400 dpi	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)

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

Sample Imaging Requirements cont'd

Works of art on paper	400 dpi	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	Uncompressed or lossless compression (e.g., LZW)
Transparencies	4000-5000 on long end or 400 dpi on output > 8" x 10"	8-bit/ channel internal reduction	Contrast stretching Minimal adjustments for tone and color	TIFF 5 & 6	uncompressed or lossless compression; (e.g., LZW)
Microfilm	600 dpi blown back to original size ---- 300-400 dpi blown back to original size	Bitonal ---- 8-bit gray	Sharpening, descreening; cropping deskewing, and despeckling	TIFF 5 & 6	Uncompressed or lossless compression (e.g., ITU-G4, LZW)

---

[2] Although 600 dpi 1-bit is a defacto 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 produces 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
- GIF only captures 256 colours

Digitization of Audio
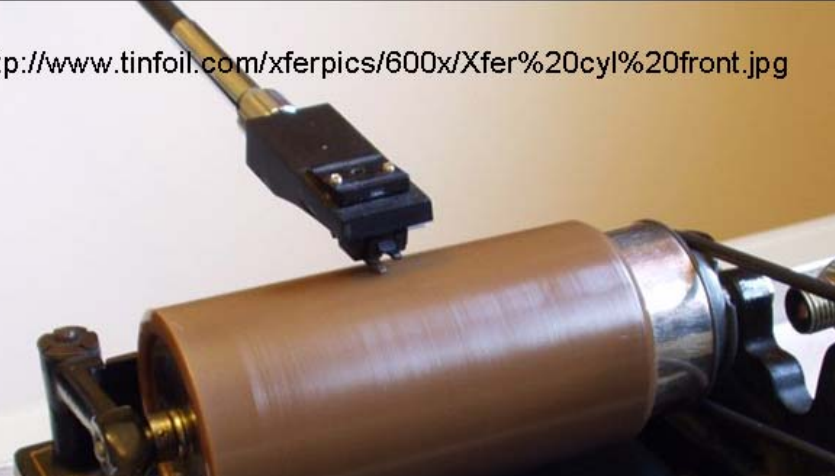
- The Digitization Process
- Audio Formats

Image:
http://www.addclasses.com/file.php/1/1earphone5-med.jpg

Wire

http://www.videointerchange.com/wire_recorder1.htm

Cartridge

http://www.videointerchange.com/audio_his

http://www.tinfoil.com/xferpics/600x/Xfer%20cyl%20front.jpg

Many formats, many devices…

nt.answers.com/main/content/img/CDE/ AUDCASS.JPG

http://www.sonicperfectionists.com/Equipment.

78 rpm shellac resin disk

http//en.wikipedia.org/wiki/Image:Vynil_record

http www.analogstereo.com/images/r2r/b77_main.jpg

Reel to reel magnetic tap

Capture Devices

Internal computer sound card

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

Image:
http://www.techexcess.net/images/products/other/sb0200_medium.jpg

Capture Devices

External analog to
digital device

- Provides
  superior results
  to sound cards

Image:
http://www.synthman.com/midiman/117212.html

Connect to ADC

Cassette players and hi-fi systems are still available can be connected to an analog to digital converter for digitization…

Direct sound output to ADC

Wire recorders, cartridge players and reel-to-reel players often have an analogue signal-out connection or can be modified by sound engineers to produce a direct sound output…

Microphone to ADC

For **wax cylinders** and other **older formats**, 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

a)

b)

c)

d)

http://www.jiscdigitalmedia.ac.uk/audio/advice/an-introduction-to-digital-audio/
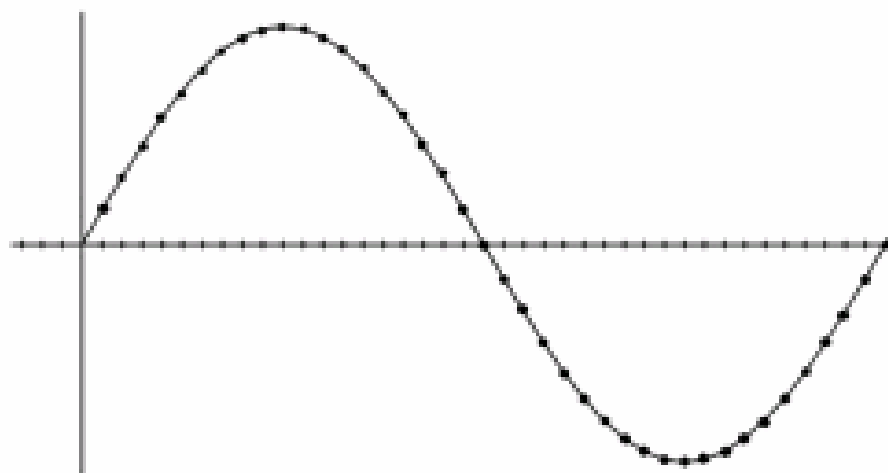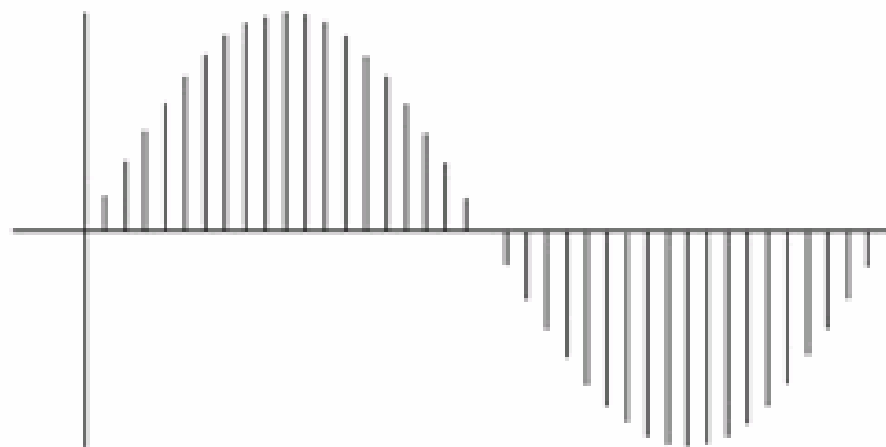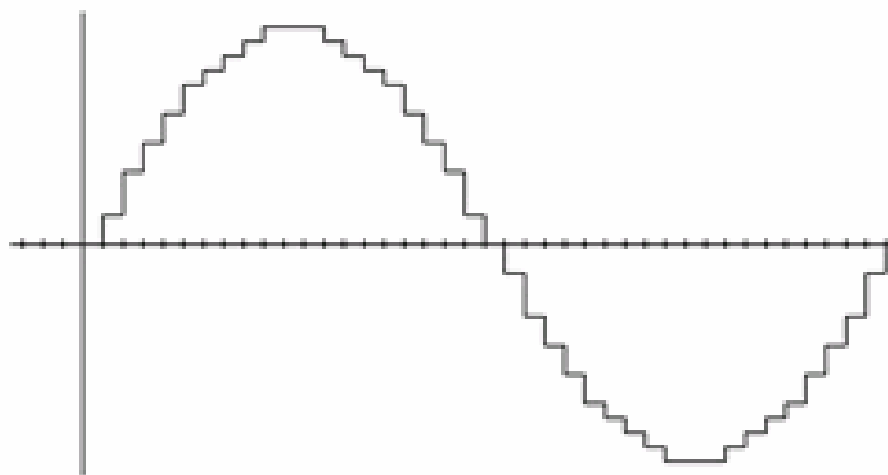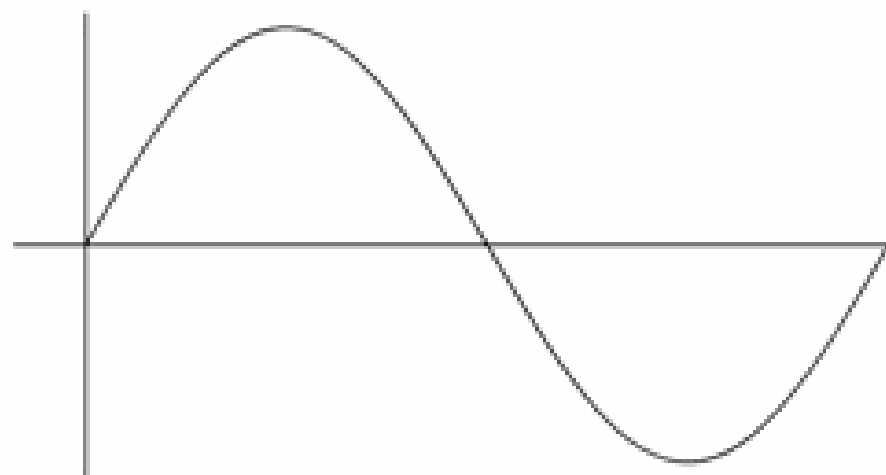
Audio Preservation Standards

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

- For more information on BWF:
  http://www.ebu.ch/en/technical/trev/trev_274-chalmers.pdf

Use and access copy

- Need expensive 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

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

Digitizing Moving Images

- The Digitization Process
- Moving Image Standard Formats
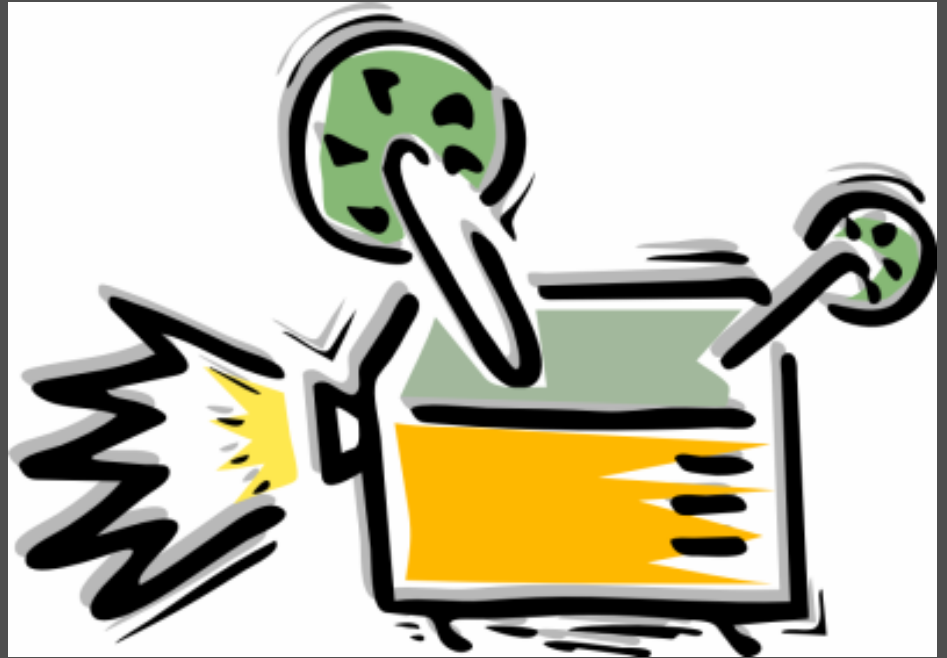
Image:
www.wpclipart.com/camera/movie_projector.png

So many different formats...

Sound
Sound
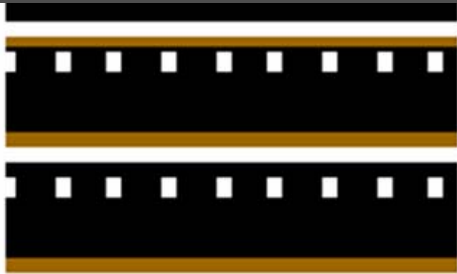
Super 8mm Film

Silen
Sour
Sour

Reel to Reel
http://www.randallareed.com/Reel_to_Reel_Tape.htm

Betacam SP

MII

8 mm video casette

http://www.lyrec.dk/images/tr532rcu.jpg
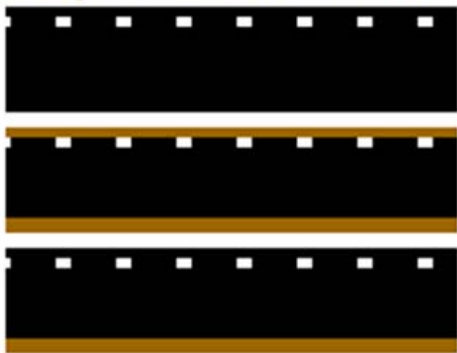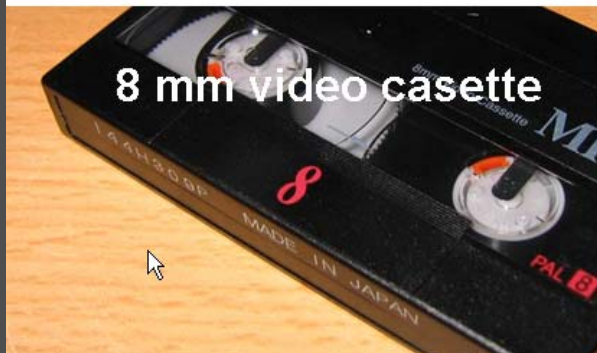
http://en.wikipedia.org/wiki/Image:Beta_tape_sizes_2.jpg
www.russellvideo.com/images/Format

Challenges

- Correctly identifying the material

- Understanding how the material was meant to be played back (eg. frame rate)

- Finding a compatible play back device:

    - In good working order

    - Within budget

    - With service professionals available

    - With extra parts available

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

WWW

The digital video files are archived onto hard drives and transcoded for distribution
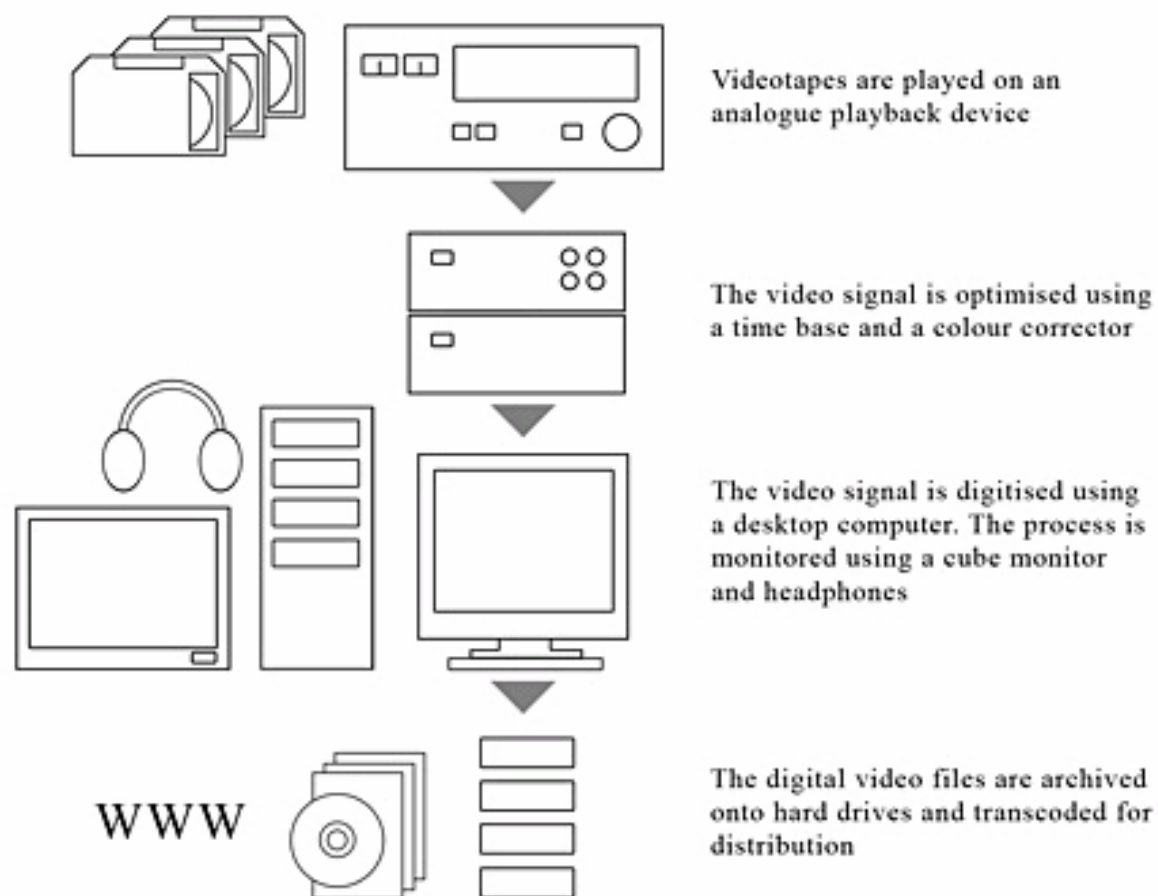
Image 1.7 Typical analogue videotape to hard drive equipment configuration (click for larger version)

http://www.jiscdigitalmedia.ac.uk/movingimages/advice/equipping-a-video-digitisation-system/

Digitizing Film

http://www.nfsa.afc.gov.au/glossary.nsf/Pages/Telecine?OpenDocument

Multiplexer (Telecine)

- Requires projector, camera, lens and mirrors

- Image projected via lens and mirrors directly into camera

- Image recorded to a common video tape format

Multiplexer (Telecine)

MULTIPLEXER OPTICAL PATH

FIRST SURFACE MIRROR

FIELD LENS

COLLIMATING LENS

T.V. CAMERA ZOOM LENS

PROJECTOR
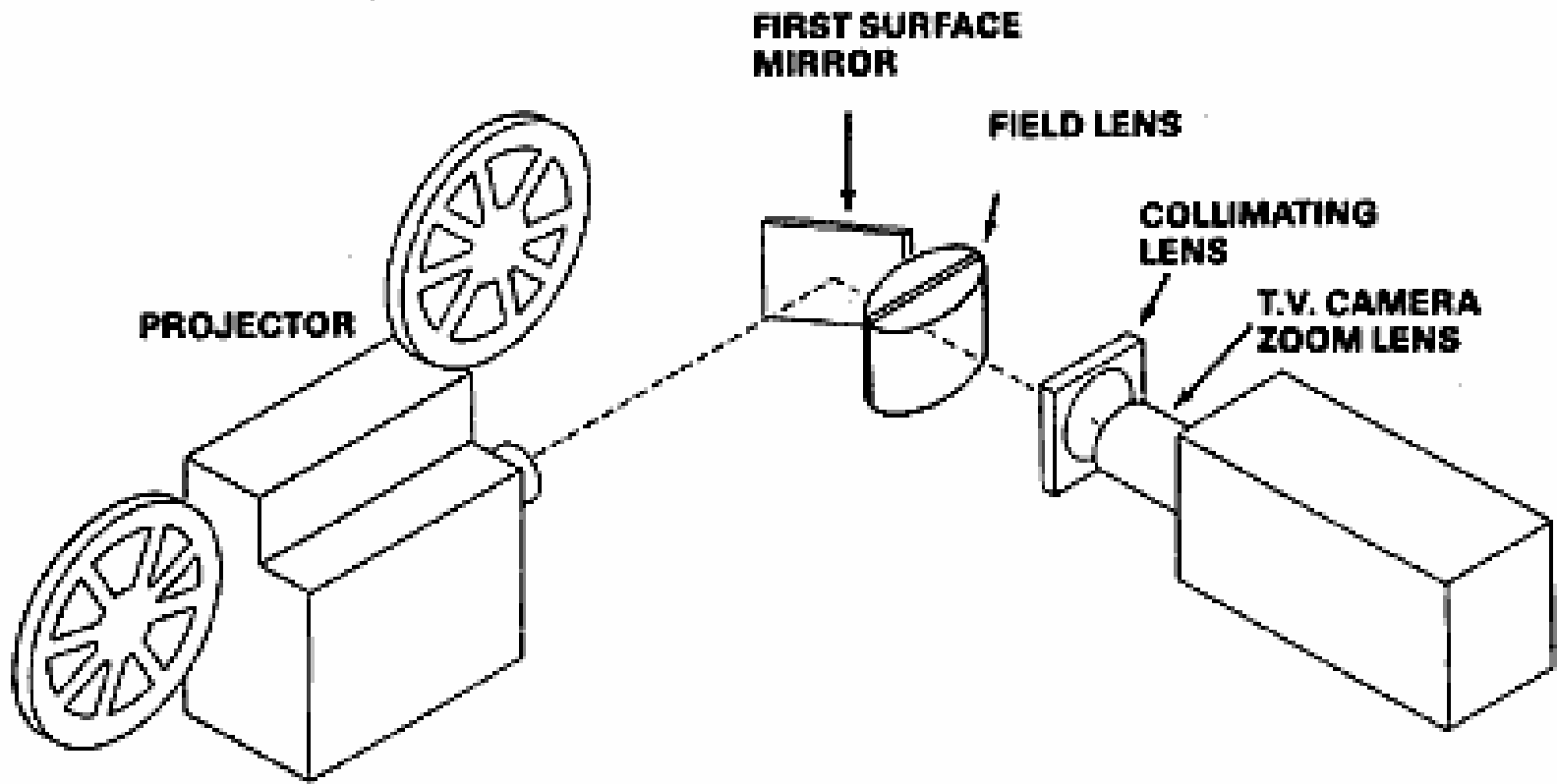
Multiplexer (Telecine)

- 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

Images:
www.visinst.com/1635Photo2.gif (top)
http://uk.gizmodo.com/flashscan8.jpg (right)

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
- Extremely expensive to acquire hardware

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

Recommendations for digital master preservation cont'd

- Higher bit rate (mb/s) preferred over lower for same lossy compression scheme

- 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

Wrapper Formats	File Formats	
	Compressed	**Uncompressed/Lossless compression**
Advanced Authoring Format (AAF)	MPEG-1	Digital Cinema Initiative Distribution Master (DCDM)
Advanced Systems Format (.asf)		
Audio Interchange File Format (.aif; .aiff)	MPEG-2	Motion JPEG (mj2, mjp2) – preferred by Library of Congress
Audio/Video Interleaved (.avi)		
Jpeg 2000 (JP2) – preferred by Library of Congress	MPEG-4	Animation codec (Quicktime)
MPEG-4	Real Video (.ram, .rm)	SheerVideo
MPEG-7	Windows Media Video format (.wmv)	
MPEG-21		
Material Exchange Format (MXF)	DivX (.divx)	
OGG format (.ogg)	Digital Video formats (DV, DVCAM, DVCPRO	
Quicktime (.mov, .moov, qt)		
Real Media (.rm)		

Format Size Comparison

Format	1 min video	1 hour video
MPEG1	10.4 MB	624 MB
WMV	12.4 MB	744 MB
AVI	214 MB	12 000 MB (12 GB)

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

Format recommendations for digital masters cont'd

High quality video (professional videotape):

- JPEG2000 uncompressed

Commercial movies:

- DCDM

Digital broadcase television streams:

- Inconclusive, industry is in a state of flux

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 loseless compression

Popular use and access formats

Streaming:

- Real Media Video

- Windows Media Video

- Quicktime

- MPEG-4 (multimedia)

Video CD:

- MPEG-1

DVD:

- MPEG-4

Where to go for help with digitization questions…

JISC — Joint Information Standards Committee

http://www.jiscdigitalmedia.ac.uk/

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

The National Archives

| About us | Visit us | Research and learning | Search the archives | Services for professionals | News | Shop online |

The **technical registry**
PRONOM

☎ **Welcome**   : **About**   ⊕ **Add an entry**
🔍 Search   ? **Help**   🔢 **Information resources**

? Help : **Simple Search**

🔍 Search : Simple Search

| Simple search | File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles | Migration Pathways |

1. Search

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

[   ]   **Search ❯**

Getting in touch

Contact us ▸

Press office ▸

Visit us ▸

Site help

Help index ▸

A-Z index ▸

Accessibility ▸

About us

Jobs and careers ▸

Terms of use ▸

Freedom of information ▸

Websites

Office of Public Sector Information ▸

Learning Curve ▸

Directgov ▸

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

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, Fedora (open source)

Kete http://kete.net.nz/

COLLINGSWOOD PUBLIC LIBRARY

The website and catalog of the Collingswood (NJ) Public Library

Hours

Monday–Friday: 9:30 a.m. to 8:30 p.m.
Saturday: 9:30 a.m. to 3:30 p.m.
Sunday: Noon to 4 p.m.

Scriblio http://about.scriblio.net/

BROWSING OUR COLLECTION

We have 60,152 books, CDs, DVDs, and other materials in our collection. You can click through the pages to see every last one of them, or click the links on the right to narrow it down.

Into The Sea
No Comments »

Brenda Z. Guiberson; illustrated by Alix Berenzy
Book, 1st, Henry Holt and Co. (BYR), 1996

Description

A beautiful, compelling book about the fragile life cycle of the endangered sea turtle.

Tags Ages 4-8 · Sea turtles · Juvenile literature · Sea turtles · Life cycles · Juvenile literature · Turtles · Juvenile literature · Guiberson, Brenda Z · Berenzy, Alix

NARROW BY SUBJECT

Animals Brothers and sisters Cats Children's poetry, American Detective and mystery stories Dogs Domestic fiction Family Fantasy fiction Friendship Historical fiction History Large type books Legal stories Love stories Police Politics and government Psychological fiction Schools Social life and customs Stories in rhyme Suspense fiction United States Women detectives World War, 1939-1945

Content DM
http://www.contentdm.com/

Browsing York University's 50th Anniversary Photograph Collection by Title

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Or enter first few letters: [_____] [Go]

Sort by: [title ▼] Order: [ascending ▼] Results: [60 ▼] [Update]

Previous Page	Now showing items 241-289 of 289

York University : Glendon College panel discussion : proposed boycott of registration discussed
Russell (1968-09-10)

York University : Glendon College panel discussion, proposed boycott of registration discussed
Russell (1968-09-10)

York University : Glendon College Women's residence opening
Unknown author (1966-10-01)

York University : Glendon College Women's residence opening
Unknown author (1966-10-01)

York University : Glendon Hall opening
Adams (1963-10-10)

Search YorkSpace

[_____] [Go]

◉ This Collection
○ Search YorkSpace

Advanced Search

Browse

- **This Collection**
  - By Issue Date
  - Authors
  - Titles
  - Subjects
  - Series
  - Fonds
  - Format
- **All of YorkSpace**
  - Communities & Collections

My Account

- Login
- Register

About YorkSpace

- What Is YorkSpace?
- Why Contribute?
- Content Guidelines
- Withdrawal Policy
- Community Policy

Opportunities for collaboration…

ADVANCED SEARCH   NEW SEARCH

HOME   GOVERNMENT DOCUMENTS | ABOUT   FAQ   HELP | EN FRANÇAIS

We found **80** matching items.
Media type(s): Video

Page 1 of 2   **1**   2   Next →

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.

Windsor Veterans' Project

Arthur V. Drake: An interview

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

Windsor Veterans' Project

Robert Murray: An interview

A video interview with Robert Murray focusing on his experiences in the RAF in World War II (edited short version)

Windsor Veterans' Project

WEB TOOLS

SHARE

Google Earth [help]

SEARCH WITHIN THESE RESULTS

Go

MEDIA TYPES

Audio (0)
Collection (0)
Image (0)
Object (0)
Text (0)
Video (80)

CONTRIBUTORS

Bibliothèque de l'Université Laval (31)
Memorial University of Newfoundland Library (31)
University of Toronto Libraries (31)
Brampton Library (24)
Windsor Veterans' Project (24)
Maritime History of the Great

http://www.ourontario.ca/

Text Archive > **Canadian Libraries**

Spotlight Item
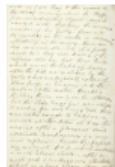
Excursion to Scotland, 1832
Seven commercially produced plates of prominent buildings and views are inserted. In addition there are a number of small hand-drawn pictures and a hand-drawn map included in the text

About the Internet Archive

Background

Frequently Asked Questions

Contributors

Alex Aitken
Alex White
Andrea Mills
Carole Moore
Dean Seeman
Gabe Juszel
Graham Stewart
John McArdle
Jonathan Bengtson
Katie Lawson
Kent Weaver
Lajolla Young

Welcome to Canadian Libraries   **184,722 items**

Historical Texts

A dictionary of the English language By Samuel Johnson(details | pdf)
A dictionary of the English language in which the words are deduced from their originals, and illustrated in their different significations by examples from the best writers : to which are prefixed, a history of the language, and an English grammar (1785)

- A general history of all voyages and travels throughout the old and new world (1708) (details | read it)
- The history of the devil, as well ancient as modern (1727) (details | read it)
- Pastorals, epistles, odes, and other original poems, with translations from Pindar, Anacreon, and Sappho (1748) (details | read it)
- Esprit, maximes et principes de m. Jean-Jacques Rousseau, de Geneve (1764) (details | read it)
- Twenty stories from Grimm (1896) (details | read it)

Photography

- Photography in the studio and in the field (1887) (details | read it)
- Beginner's guide to photography (1888) (details | read it)
- Uncle Alberts manual of practical photography (1890) (details | read it)
- Studio light ... a magazine (1910) (details | read it)
- The evolution of photography (1890) (details | read it)

Ephemeral Canadiana

- Tremendous Toronto (1914) (details | read it)
- Se-ke and the beaver (1920) (details | read it)
- Grimsby Park, historical and descriptive: with biographical sketches of the late president Noah Phelps and others (1900)

Most Downloaded Items Last Week more

1. Who's who in China; containing the pictures and biographies of China's best known political, financial, business and professional men
   1,746 downloads

2. 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...
   1,658 downloads

3. A dictionary of the English language : in which the words are deduced from their originals, and illustrated in their different significations by examples from the best writers : to which are prefixed, a

http://www.archive.org/details/toronto

MIC collects 558,489 catalog records from 15 participating institutions.
http://mic.loc.gov/

Metadata

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

```
<titlStmt>
<titl>Survey for Rural Economies, 1998</titl>
<IDNo>7856</IDNo>
</titlStmt>
<fundAg>Countryside Agency</fundAg>
<fundAg>Department of the Environment, Transport and the
Regions</fundAg>
<copyright> Social Research Centre</copyright>
<grantNo>3289460</grantNo>
<distStmt>
<distrbtr abbr="UKDA" affiliation="University of Essex, Wivenhoe Park,
Colchester, Essex, England, CO4 3SQ">UK Data Archive</distrbtr>
<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

http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

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:

Title	Source	Contributor
Creator	Language	Date
Subject	Relation	Type
Description	Coverage	Format
Publisher	Rights	Identifier

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

Agnew, G., Kniesner, D., & Weber, M. B. (2007). Integrating MPEG-7 into the moving image collections portal. *Journal of the American Society for Information Science and Technology, 58*(9), 1357-1363.

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

Recommended mimium 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:
http://www.jisc.ac.uk/media/documents/programmes/
preservation/moving_images_and_sound_archiving_
study1.pdf

Browse the Collection

Author

View highlights from the Collection

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 >>

START/STOP

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

870   1890   1900
Interactive Timeline ▷

Interactive Map ▷

Educational Toolkit ▷

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

Author ▾

Browse by Field
- **Author**
- Title
- Subject

Browse by Media
- Video
- Audio
- Photographs
- Drawings/Paintings
- Plans/Maps
- Manuscripts
- Published Texts

Browse by Theme
- Geography
- Architecture
- Education
- Travel
- Language
- Lifestyle and Customs
- Commerce and Trade

Author ▾

- Technology
- Religion
- Health
- Music

Browse by Community
- Ramah
- Okak
- Nain
- Hopedale
- Makkovik
- Hebron
- Killinek
- Zoar
- Ailik
- Uviluktok
- Eclipse Harbour

Browse by Language
- French
- Inuktitut
- German

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 h)

- 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.htm

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
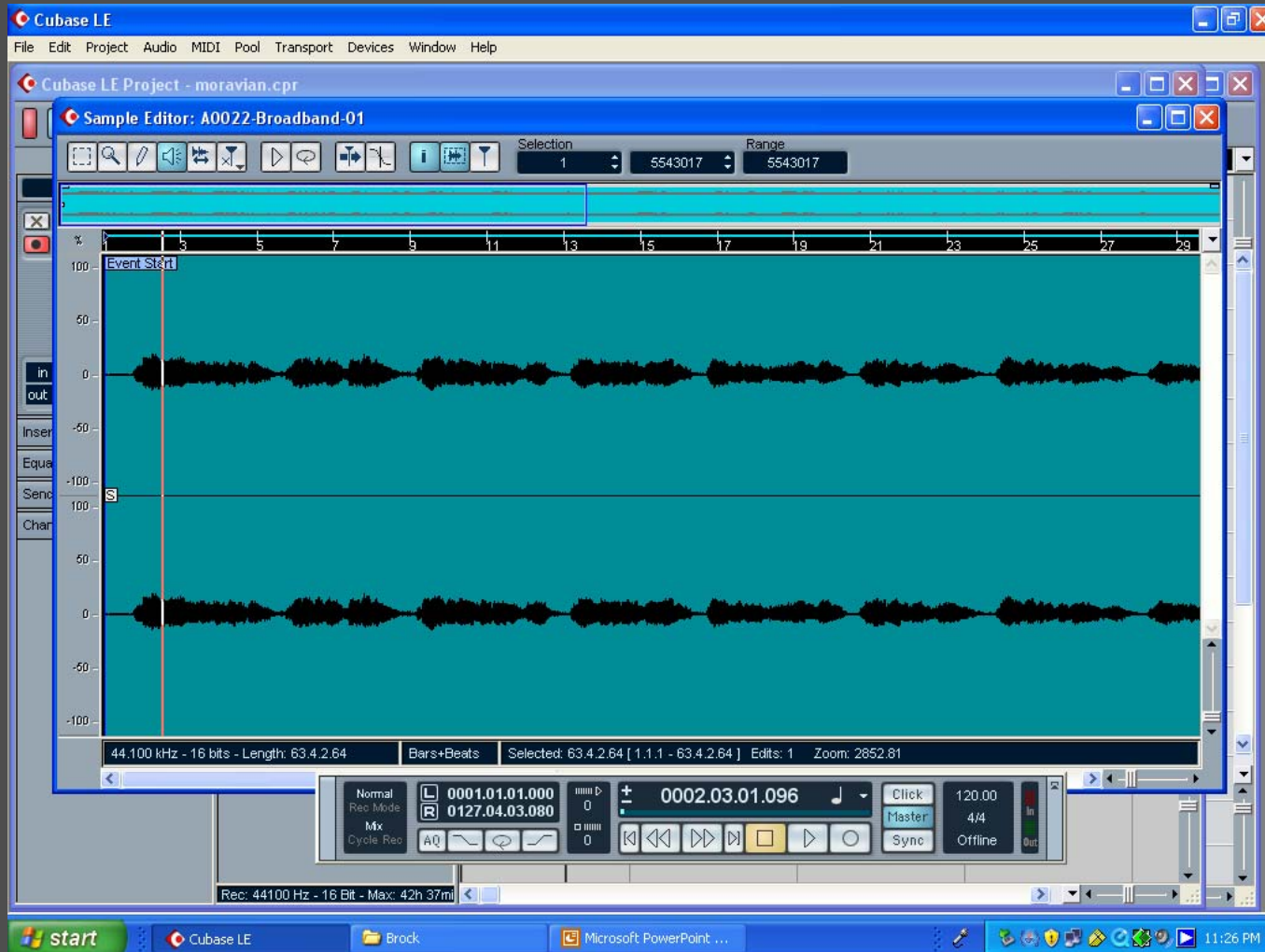
Basic Sound Recording Principles

- 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
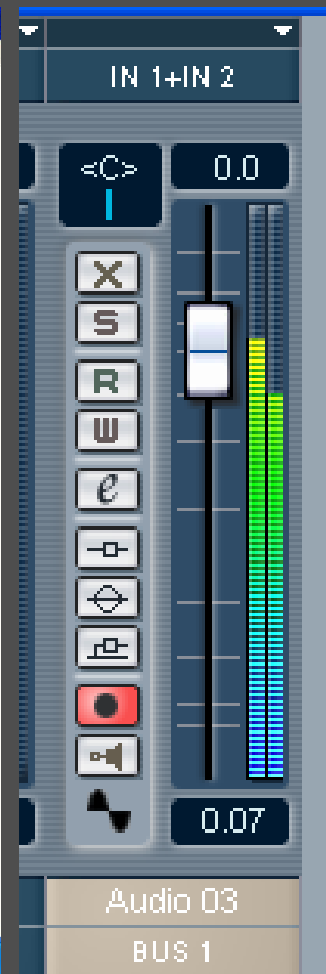- We captured files too quietly, had to go back and boost levels

Example of a clipped wave
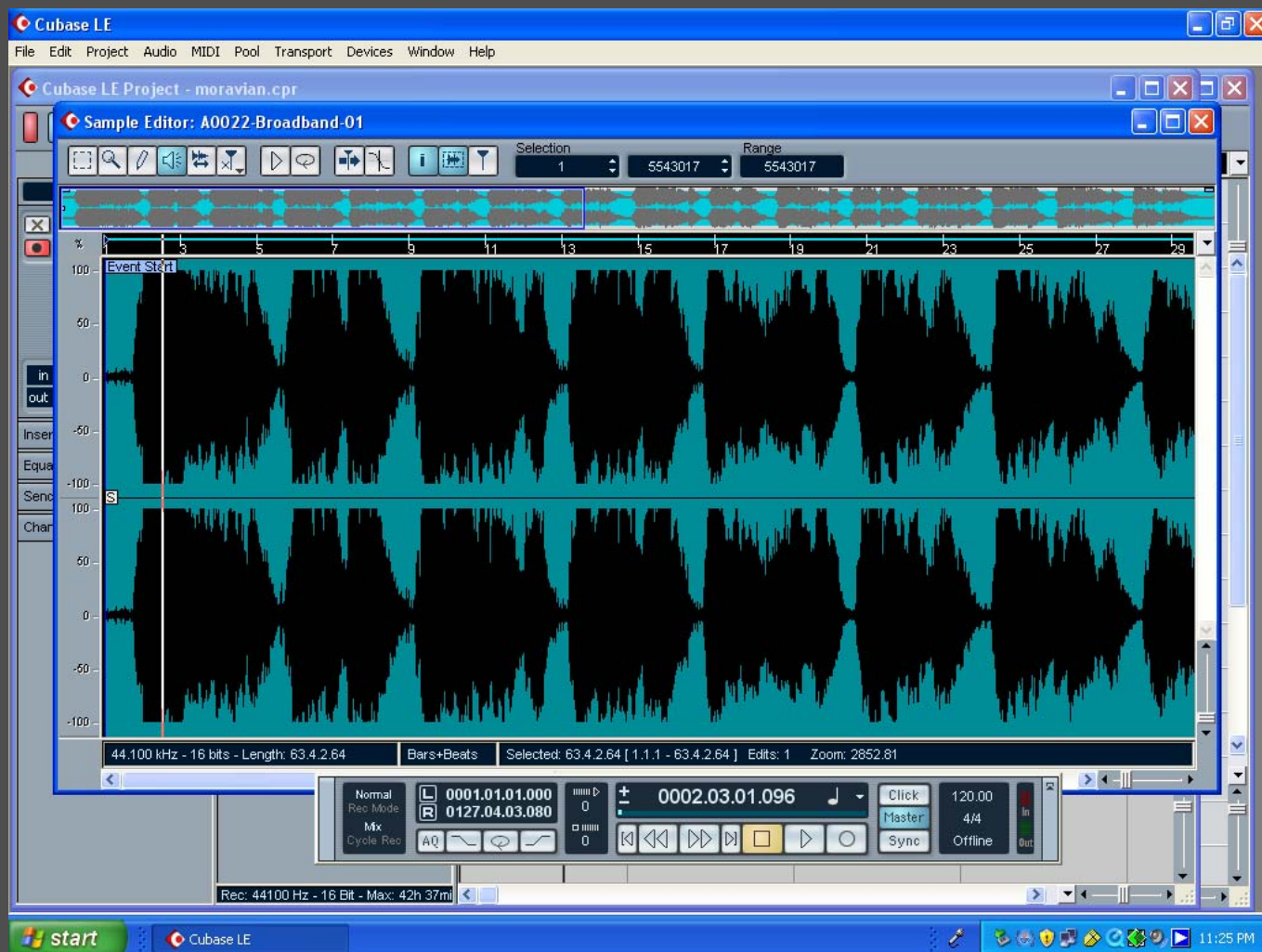
Example of a wave that needs boosting

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)

Happy Digitizing!