WORKING WITH CONTENT CREATORS

June 27, 2014

Yvonne Ng
Senior Archivist, WITNESS
yvonne@witness.org | @ng_yvonne
WHERE WE ARE IN THE WORKFLOW

OAIS

[Diagram showing the workflow process involving Preservation Planning, Ingest, Archival Storage, Access, and Administration.]
PAIMAS

• Preliminary phase
• Formal Definition phase
• Transfer phase
• Validation phase
WHY SUPPORT CONTENT CREATORS?
WHAT IS IT?
INCOMPLETE PACKAGES
archiveguide.witness.org

Start Here
- Who is this Guide for?
- Why Archive?
- What is Archiving?
- How to Use this Guide

The Workflow
- Create
- Transfer
- Acquire
- Organize
- Store
- Catalog
- Preserve
- Share

Resources
- Key Concepts
- Glossary
- Tip Sheets
- Video as Evidence

Create
Start archiving your video at the point of creation.

Transfer
Move video and metadata from one device or location to another.

Acquire
Receive video and metadata from a source and add it to your collection.

Organize
Organize your digital video files and documentation.

Catalog
Create a structured and searchable system for creating and managing metadata, and finding your videos.

Preserve
Ensure that your collection will be preserved and accessible far into the future.

Share
Make your collection accessible to others outside your organization.
Organize: Filenames

Do Not Rename Camera Filenames

A filename generated by a camera provides valuable information about a video. For example, in a group of video files, you can easily tell that DSC_991.AVI was recorded after DSC_990.AVI, but before DSC_992.AVI, and that it was not recorded on the same camera as VID00005.AVI.

Retaining the filename is part of maintaining the original order, which is important for evidence and contextualization. In addition, some complex video formats rely on the original filename to function properly.

```
| CLIPINF | 00014.MTS |
| index.bdm | 00015.MTS |
| MOVIEOBJ.BDM | 00016.MTS |
| PLAYLIST | 00017.MTS |
| STREAM | 00018.MTS |
|          | 00019.MTS |
|          | 00020.MTS |
|          | 00021.MTS |
```

Raw camera footage in its original order.

What If I Have Duplicate Filenames?

Duplicate filenames can sometimes occur when you acquire raw footage files from different people who use the same camera brand. Rather than renaming the files, organize footage from each source person into a separate folder (see “Folders and Directories” for more on this).
Organize: Using Unique Identifiers

What is a Unique Identifier?

A unique identifier is a number or code that can unambiguously distinguish one object from another in a given system, and group things associated with an object together. We frequently encounter unique identifiers in our everyday lives, such as credit card numbers, phone numbers, barcodes, and book ISBNs. A credit card number, for instance, distinguishes your purchases from someone else’s, and allows all of your purchases to be grouped together on one bill.

A product barcode is an example of a unique identifier.

You can use unique identifiers to organize your videos. Imagine that you have 10 video files acquired from various sources, whose camera-assigned filenames have all been changed. When you review the videos, you find that 9 are unique, and one is a copy. You can create 9 unique identifiers to distinguish the 9 videos from one another. You can also give the one non-unique copy the same unique identifier as its original to associate them together.
Organize: Folders and Directories

Original Order

The key principle when organizing or grouping raw video footage into folders is to preserve its original order. In the context of video documentation, the original order is the order in which video files are recorded— for example VID00001.avi, VID00002.avi, VID00003.avi. The original order of files has evidential significance; you can infer that VID00002.avi was shot after VID00001.avi, and nothing was filmed in between. Original order also provides context to the individual files -- for example, the events depicted in VID00003.avi somehow relate to the events in VID00002.avi and VID00004.avi.

Note that you may not always receive files in their original order, and that you may need to restore the original order when you organize your files.

Folders as Information Packages

A simple way to organize your videos and documentation is to make information packages. “Information package” is a term borrowed from the archiving world that simply refers to a container, such as a folder, that holds the object being archived— your videos, photos, recordings, notes, etc.
Organize: Tools for Media Management

Using your Finder or File Explorer to browse organized directories and folders may be perfectly sufficient for your needs. However, if you need to navigate or sort your files in more complex ways there are many tools ranging from simple to highly sophisticated that you can use alongside your organized directories.

Personal Media Management Applications

Personal media management tools often come pre-installed on your computer, or can be purchased for a low cost. These systems are usually very easy to use, but are limited in their functionality, so are best suited to small collections. Some examples include:

- iTunes
- iPhoto
- Windows Media Center

Note that personal media management tools are not usually built to allow you to export your information to other systems. If you enter a description of a video in iTunes library, for example, you cannot easily move that information into another system later on.

Video Production Media Management Systems

Media management systems built for video production usually offer extra functionality beyond organizing your videos, which you may or may not need, such as logging, transcoding, batch processing tools and integration with video editing systems. These systems also usually provide more access to your video’s technical metadata. Some examples include:

- CatDV
- Adobe Bridge
Start Here: What is Archiving?

Archiving is... a general term for the range of practices and decisions that support the long-term preservation, use, and accessibility of content with enduring value. In this Guide, our focus is on your digital videos.

Archiving is... an ongoing process that begins when a video is created and continues infinitely into the future.

Archiving is...a process that can be incorporated into your existing video workflows.

Archiving is... a way to ensure your videos remain authentic and intact, so you can use them as evidence.

Archiving is... a way to ensure your videos are available, findable and playable long into the future.

Archiving is NOT... a one-time action.

Archiving is NOT... putting your videos on a hard drive and leaving it on a shelf.
A practical resource to help you manage, authenticate, and preserve your digital video.

**Create**
Start archiving your video at the point of creation.

**Transfer**
Move video and metadata from one device or location to another.

**Acquire**
Receive video and metadata from a source and add it to your collection.

**Organize**
Organize your digital video files and documentation.

**Store**
Store your videos, manage the storage environment, and recover from errors.

**Catalog**
Create a structured and searchable system for creating and managing metadata, and finding your videos.

**Preserve**
Ensure that your collection will be preserved and accessible far into the future.

**Share**
Make your collection accessible to others outside your organization.
Create: What Metadata to Capture

Key contextual information about your video needs to be captured at the time it is created. This metadata is critical to the video’s authenticity, and to the ability to find, use, and understand the video.

The key pieces of information to capture at the point of creation are:

» **When**

The date and time recorded / created.

» **Where**

The geographic location of recording.

» **What and why**

A basic description - the important details about the event recorded that would be difficult to identify later (e.g. people’s names, the purpose for recording) or that make the event significant (e.g. shelling during a ceasefire, violence against civilians).

» **Who**

The video’s source. The full name (or pseudonym, if not safe) and contact information (if safe to provide) of the video’s creator.

» **Security requirements**

Whether or not the identities of the video’s subjects or creator need to be protected.

Other information, such as detailed descriptions or keywords, will be important to making your video more findable and understandable but are not critical at this stage. Additional information can be added later at different stages of the workflow (see “Catalog”).
CREATE
TRANSFER
### Original Transfer Details

**Video File:** IMG_2176.MOV

- **Format:** MPEG-4
- **Format profile:** QuickTime
- **Codec ID:** Qt
- **File size:** 965 KiB
- **Duration:** 11s 872ms
- **Overall bit rate:** 666 Kbps
- **Recorded date:** 2012-10-26T17:24:33-0400
- **Encoded date:** UTC 2012-10-26 21:25:30
- **Tagged date:** UTC 2012-10-26 21:25:35
- **Writing application:** Apple QuickTime
- **Writing library:** Apple
- **Make:** Apple
- **OXYZ:** +40.6851-073.9742+040.776/
- **Model:** iPhone 4S

- **com.apple.quicktime.make:** Apple
- **com.apple.quicktime.creationdate:** 2012-10-26T17:24:33-0400
- **com.apple.quicktime.location.ISO6709:** +40.6851-073.9742+040.776/
- **com.apple.quicktime.software:** 6.0
- **com.apple.quicktime.model:** iPhone 4S

### Transcoded Video Details

- **File Name:** /Users/yvonne/Desktop/IMG_2176.mp4
- **Format:** MPEG-4
- **Format profile:** Base Media / Version 2
- **Codec ID:** mp42
- **File size:** 520 KiB
- **Duration:** 11s 900ms
- **Overall bit rate:** 358 Kbps
- **Encoded date:** UTC 2012-10-29 14:38:36
- **Tagged date:** UTC 2012-10-29 14:38:36
- **gsst:** 0
- **gstd:** 12026
- **gsdd:** B4A7DD601MM1351608287635719
- **gshh:** r1---sn-pSusan7e.c.youtube.com
ACQUIRE
STORE

copies

Onsite Backup
Primary Storage

Offsite Backup

access control

fixity check

refresh
STORE
CATALOG

<table>
<thead>
<tr>
<th>DATE RECORDED</th>
<th>SUBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-05-01</td>
<td>protest</td>
</tr>
<tr>
<td>1-May-13</td>
<td>protest</td>
</tr>
<tr>
<td>05/01/2013</td>
<td>demonstration</td>
</tr>
<tr>
<td>May 1, 2013</td>
<td>protest</td>
</tr>
<tr>
<td></td>
<td>protest</td>
</tr>
</tbody>
</table>
Catalog: Getting Started

Start with an Inventory

Building a catalog can be a labor-intensive process, and cataloging requires at least some training. Start small by creating a simple inventory of your collection. An inventory is a list of your videos with only essential information such as ID, file or folder name, title, storage location(s), and security restrictions. An inventory can provide basic access to your collection until you are able to build a more elaborate catalog, and the data you enter in the inventory can be incorporated into it.

Example of a typical inventory.

Evaluate your Collection

Before deciding to build a catalog, evaluate whether you really need something more complex than an inventory (and acquired metadata and documentation) to make your videos findable and understandable. An inventory may be sufficient.

If you do want to make a catalog, consider aspects of your collection that will affect the complexity of the catalog and time required to build it:

» Collection size

How many items do you have? What is your rate of acquisition? The catalog you need for 100,000 titles will be different from one for 10,000.

» Information needs

How complex is the information you need to capture?

» Access needs

How do you or your current or future users need to access information? What are the gaps in your current ability to search, and what functionality does your catalog need to have to address them? What are the critical access points?
Choosing an Archive

There may be one or many institutions or organizations interested in acquiring your collection. When choosing a potential archive, there are several factors you should consider:

» Trustworthiness

Do you trust the archive (and the institution it may belong to) to take care of your collection and abide by its agreements with you (e.g. regarding security restrictions, access, preservation)?

» Resources

Does the archive have the staffing and infrastructure to meet the processing, storage, preservation, and access needs of your collection?

» Collecting focus

Does the archive have a real interest in your collection, and experience and expertise in dealing with collections similar to yours?

» Restrictions/ access

Can the archive accommodate your expectations for security and privacy restrictions?

» Ownership

Do you want to retain ownership of your collection, or are you willing to transfer ownership to the archive? Some archives will accept collections they do not own, but some will not.

» Rights

Do you own the copyright or have rights to the content in your collection? If not, can you provide the archive with information about third-party rightsholders? Archives need to understand the rights restrictions in order to provide access.

» Deposit logistics

Are you able to get your collection to the archive?

Donor/Deposit Agreements

If you work with an archive, draft a written agreement that outlines their acquisition of your collection and the terms of your relationship. This ensures that both sides clearly understand their rights and obligations, which ultimately protects the collection.
Preserve: Other Preservation Options

You may have reasons for not wanting to deposit your collection at an established archival institution. Besides working with an archive, two other available options are to establish your own archive, and/or deposit your collection with a unique non-profit called the Internet Archive.

Establishing Your Own Archive

It is challenging, but not impossible, to establish and sustain an archive on your own or with a network of like-minded organizations. This option requires significant ongoing infrastructure, human resources, and financial support.

Establishing an archive involves building a repository, developing archive policies and procedures, and performing the day-to-day work of acquiring, cataloging, preserving, and providing access to users.

Try This Advanced

CAIS Reference Model is an International Organization for Standardization (ISO) standard that defines archive concepts and establishes the minimum requirements for an archive.

Audit and Certification of Trustworthy Digital Repositories provides metrics for measuring the trustworthiness of your digital repository.

Internet Archive-backed Archive

The Internet Archive is a unique non-profit digital library that allows the public to upload and download digital material at no charge. Its mission is to provide permanent access to historical content in digital format. The Internet Archive holds
Start Here

Who is this Guide for?
Why Archive?
What is Archiving?
How to Use this Guide

The Workflow

- Create
- Transfer
- Acquire
- Organize
- Store
- Catalog
- Preserve
- Share

Resources

- Key Concepts
- Glossary
- Tip Sheets
- Video as Evidence

Create
Start archiving your video at the point of creation.

Transfer
Move video and metadata from one device or location to another.

Acquire
Receive video and metadata from a source and add it to your collection.

Organize
Organize your digital video files and documentation.

Catalog
Create a structured and searchable system for creating and managing metadata, and finding your videos.

Preserve
Ensure that your collection will be preserved and accessible far into the future.

Share
Make your collection accessible to others outside your organization.
Create a structured and searchable system for creating and managing metadata, and finding your videos.

Ensure that your collection will be preserved and accessible far into the future.

Make your collection accessible to others outside your organization.
TOOL: RSYNC
RSYNC

- command line utility for transferring / copying files
- commonly used for backups
- has many options, allowing you to control behavior and specify how files are copied
HOW WE USE IT

- To copy/move files to archive storage
- To write backups to LTO (as part of the “writeLTO” script available here: https://github.com/amiaopensource/ltopers
WHY WE USE IT

- copies efficiently, can skip files, can restart easily if interrupted
- uses rolling checksums and checksum at end of file transfer to ensure proper transfer
- provides us with confirmation/logs
- can preserve file attributes like modification times
- can be used in other programs/scripts
BASIC RSYNC SYNTAX

• e.g. Copying a directory from my Desktop to mounted archive storage volume, preserve modification times, print information about execution of command.

Yvonne-MB:~ yvonne$ rsync -rtv /Users/yvonne/Desktop/Photos /Volumes/SAN1

command   options   source   destination
TOOL: MD5 & MD5DEEP
MD5

- command line utility for generating MD5 message digests ("hashes")

- MD5 is an algorithm that generates a 128-bit hash, usually represented in 32 hexadecimal digits, e.g. 61b7ef0c6f99cc4dfb03b418419be545
MD5DEEP

- Set of programs that can generate and check MD5, SHA-1, SHA-256 and other message digests.
- Can generate and check hashes recursively on files in a directory.
HOW WE USE IT

• Ad hoc, to quickly compare two files/folders to see if they are the same.

• To monitor stored data objects for fixity over time (using a program/script that runs MD5 on all archived files on a regular basis).
WHY WE USE IT

• easily check if files transferred properly

• a way to monitor large number of stored files for fixity over time

• you may want to try AVPS Fixity Tool (free!): http://www.avpreserve.com/avpsresources/tools/
BASIC MD5 SYNTAX

- e.g. Generate hash for two files

```
MD5 (/Users/yvonne/kSXVJWN-Bqs.flv) = f09c1ff7ffa3818171d5c5a5de239a12
MD5 (/Users/yvonne/k98JcvfPFrU.mp4) = 3f9588472c2e7883a74a5276a2621efa
```
BASIC MD5DEEP SYNTAX

• e.g. Generate hashes for file directory and save as text file

Yvonne-MB:~ yvonne$ md5deep -r /Users/yvonne/Desktop/Photos > my_hashes.txt

• Check files against list of hashes for FILES THAT MATCH LISTED HASH

command ↓ “matches this list” ↓ recursive ↓ input files

Yvonne-MB:~ yvonne$ md5deep -M /Users/yvonne/my_hashes.txt -r /Users/yvonne/Desktop/Photos
194577a7e20bdcc7afbb718f502c134c /Users/yvonne/Desktop/Photos/.DS_Store
cd46cada8ebb9d47ca22228ce41c5b97 /Users/yvonne/Desktop/Photos/IMG_0990.JPG
f181290084d71c4f348b5d072fa90ba /Users/yvonne/Desktop/Photos/IMG_0991.JPG
51ee8ffbabac71965d473b359d3237f1 /Users/yvonne/Desktop/Photos/IMG_1025.JPG
BASIC MD5DEEP SYNTAX

• Check files against list of hashes for FILES THAT DO NOT MATCH LISTED HASH. (I added HAHA.gif)

Yvonne-MB:~ yvonne$ md5deep -X /Users/yvonne/my_hashes.txt -r /Users/yvonne/Desktop/Photos 75f4ea124b73bf876ba3a2815e2056f7 /Users/yvonne/Desktop/Photos/HAHA.gif

• Check file directory against list of hashes for HASHES WITH NO MATCHING FILES. (I deleted a file from my directory)

TOOL: BAGIT
BAG IT

• file packaging “format” for transferring and storing files.

• basically, a standardized file directory structure with certain required documents inside:
BAGIT .TXT ELEMENTS

BagIt-Version: M.N
Tag-File-Character-Encoding: UTF-8

cd46cada8ebb9d47ca22228ce41c5b97       data/IMG_0990.JPG

ace0ef9419c8edbe164a888d4e4ab7ee
3f638ffada6855727f86ad3833f7ce03
cde041ff2d5b0a72dd73dade14d555b4

Source-Organization: Spengler University
Organization-Address: 1400 Elm St., Cupertino, California, 95014
Contact-Name: Edna Janssen
Contact-Phone: +1 408-555-1212
Contact-Email: ej@spengler.edu
External-Description: JPG images from WITNESS
Bagging-Date: 2014-06-24
External-Identifier: spengler yoshimuri_001
Bag-Size: 2.8 MB
Payload-Oxum: 279164409832.1198
Bag-Group-Identifier: spengler yoshimuri
Bag-Count: 1 of 1
Internal-Sender-Identifier: /storage/images/
Internal-Sender-Description: picture of Yvonne
HOW WE USE IT

• To package and send video masters to University of Texas Libraries for long-term preservation.

• To store video masters in University of Texas repository.
WHY WE USE IT

• we didn’t have to think of it ourselves

• easily package video files along with XML catalog records and other documentation

• available tools for creation and validation

• ensures that what we sent is what is received
TOOL: BAGGER
BAGGER

- graphical user interface for creating and validating Bags
- built on the Bag It Library, a java software library
This space will contain messages generated by the creating and updating of bags. Mouse over the status label to see its description.

[Tue Jun 24 16:48:03 EDT 2014]: A new bag has been created in memory.