



Paralysis often strikes when we are unable to quantify and articulate, and this has been the case for many organizations when it comes to working with audiovisual and digital preservation. Today we're going to talk about free and open resources that all help organizations enable action and move away from paralysis or a state of stagnation. We will hit at three different levels today. First Yvonne Ng from WITNESS will talk about The Activist's Guide to Archiving Video. An amazing resource aimed at activists but relevant to everyone concerned with video archiving. The guide speaks to workflows, processes and practices. Erin Engle from the Library of Congress is going to talk about the NDSA Levels of Digital Preservation. An effective framework for thinking about and assessing digital preservation infrastructure. I will be talking at a lower level about two specific tools.



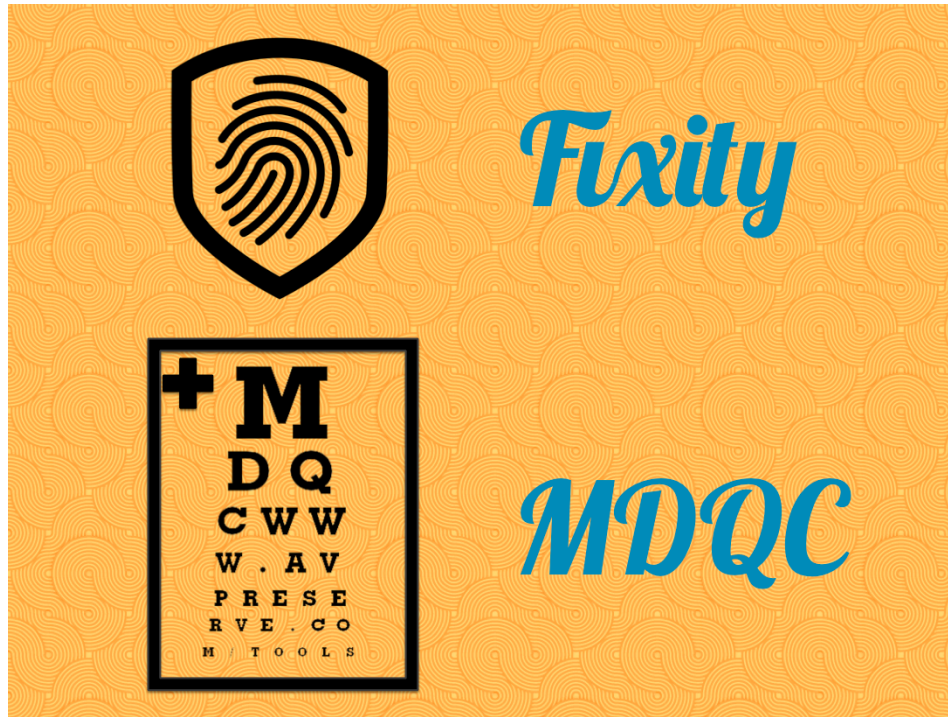
I'll start off by introducing Yvonne, Erin and myself up front, and we will take questions at the very end of all three talks.

Yvonne Ng is the Archivist at WITNESS, an non-profit organization that supports human rights activists to use video to document abuses and advocate for change. Prior to joining WITNESS in 2009, Yvonne was a Research Fellow on the NDIIPP Preserving Digital Public Television project. Yvonne has also worked at New York University Libraries, New York Public Library, and the Canadian Filmmakers Distribution Centre. She is a graduate of the NYU Moving Image Archiving and Preservation program.

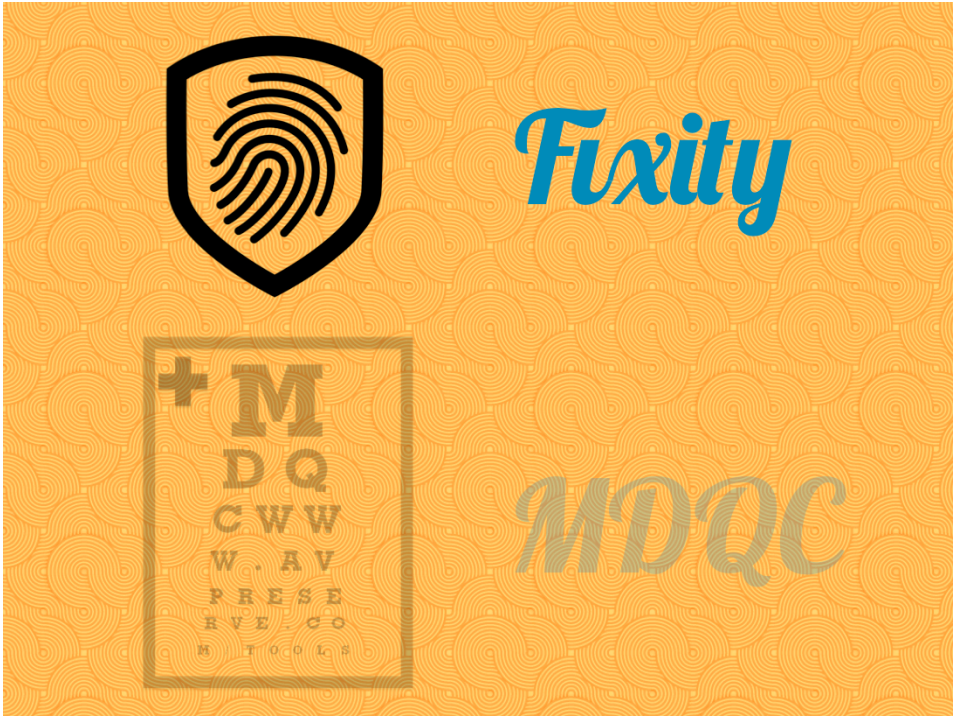
Erin Engle is a Digital Archivist with the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress. She coordinates and contributes to communications, outreach, and social media activities for the Program and is the communications officer for the National Digital Stewardship Alliance. She has a BA in International Politics and French from The Pennsylvania State University and she received her MSLS from The Catholic University of America.



I am Chris Lacinak from AVPreserve. We are a consulting and software development firm that works with organizations on activities ranging from collection assessments to digital asset management and preservation. Our clients include the Library of Congress, Indiana University, NYPL, The New York Six Consortium and the Museum of Modern Art. Aside from doing work for hire we also develop lots of resources including free and open source tools like the ones I'll be talking about today.



The two tools I'll be talking about today are Fixity and MDQC.



Starting off with Fixity...

WHAT DOES IT DO?



Checksum Generation and Validation
MD5 or SHA256

WHAT DOES IT DO?



File Attendance
Filepath
inode

WHAT DOES IT DO?

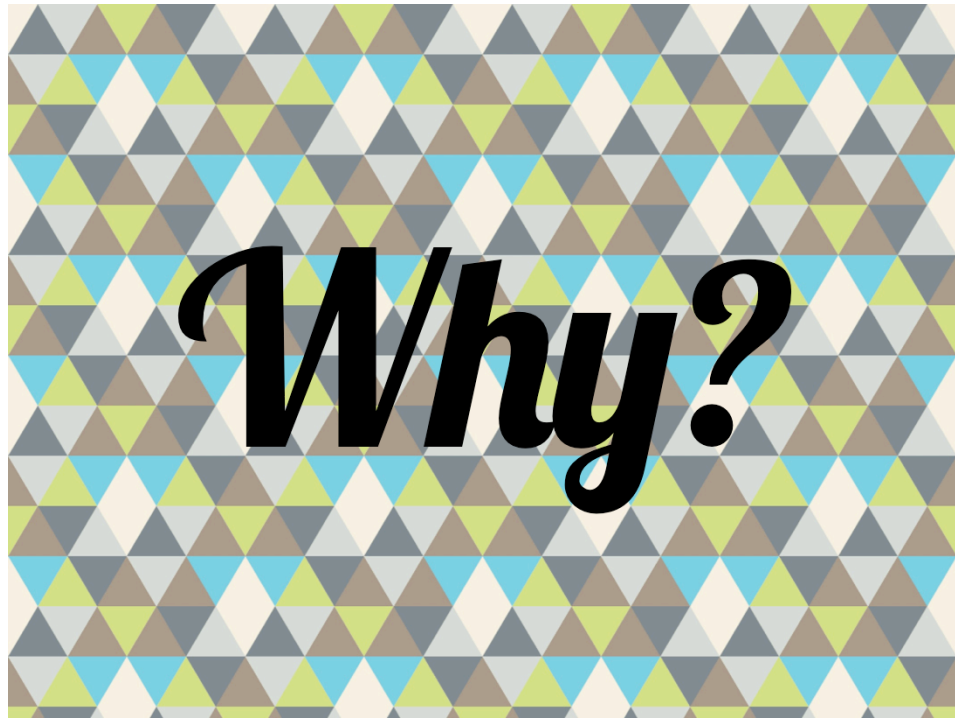


Automated scanning and reporting according to a user-defined schedule

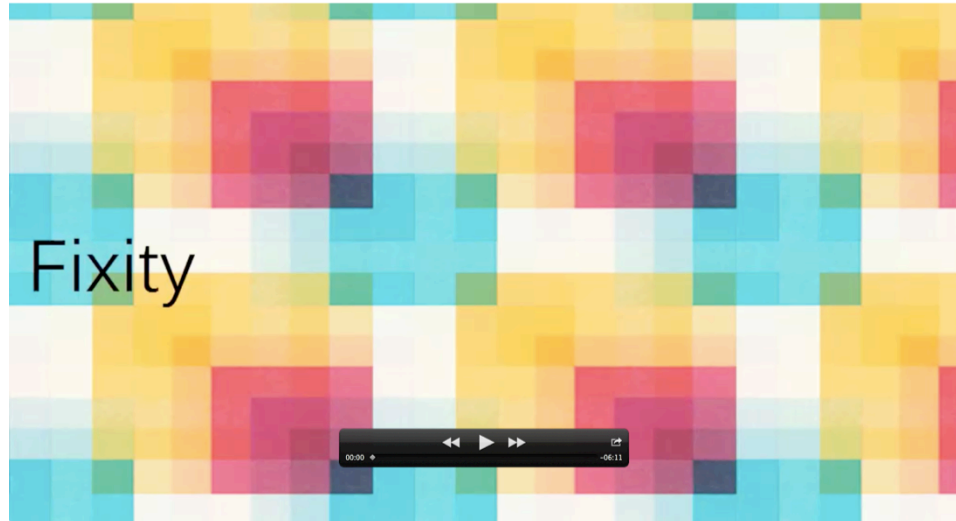
WHAT DOES IT DO?



Reporting locally and to email



- Fixity is one of these things that EVERYONE talks about, and some even generate checksums but VERY FEW have routine validation in place
- To do this using current toolset requires IT expertise that is beyond most organizations
- All other MD5 apps are for other workflows
- No other tools with similar functionality that were accessible to non-IT experts
- No other tool incorporates ability to identify moved/renamed files or file attendance and integrity



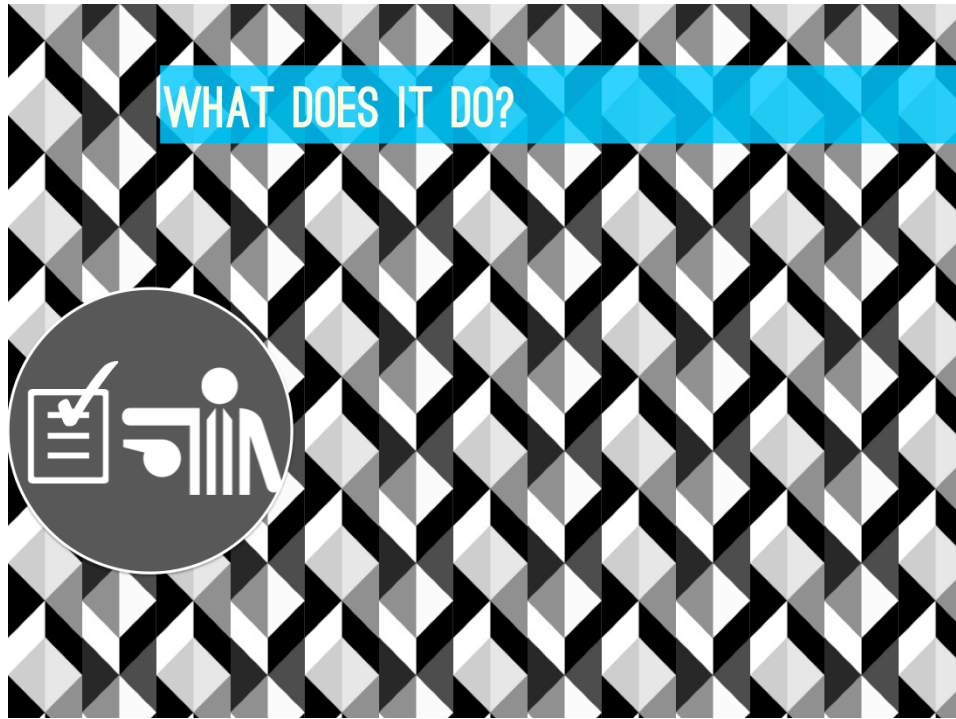
See S12-Lacinak_fixity_demo.mp4



Fixity

+ M
D Q
C W W
W . A V
P R E S E
R V E . C O
M / T O O L S

MDQC



Uses a reference file in order to help establish the parameters and rule set for comparison
Define ruleset

EXAMPLES OF EMBEDDED METADATA

Statement of Work Sample.docx Properties

General Summary Statistics Contents Custom

Title: Sample Statement of Work

Subject: Outsourcing Digitization

Author: Chris Lacinak

Manager: Chris Lacinak

Company: AVPS

Category:

Keywords: Sample, Statement of Work, Digitization, Video

Comments: Used as a reference for demonstrating an appropriate level of detail to provide to vendors when outsourcing reformatting.

Hyperlink base: www.avpreserve.com

Template: Normal.dotm

Save preview picture with this document

Cancel OK

EXAMPLES OF EMBEDDED METADATA



This is what we see in the photo, but when we look at the embedded metadata we see....

EXAMPLES OF EMBEDDED METADATA

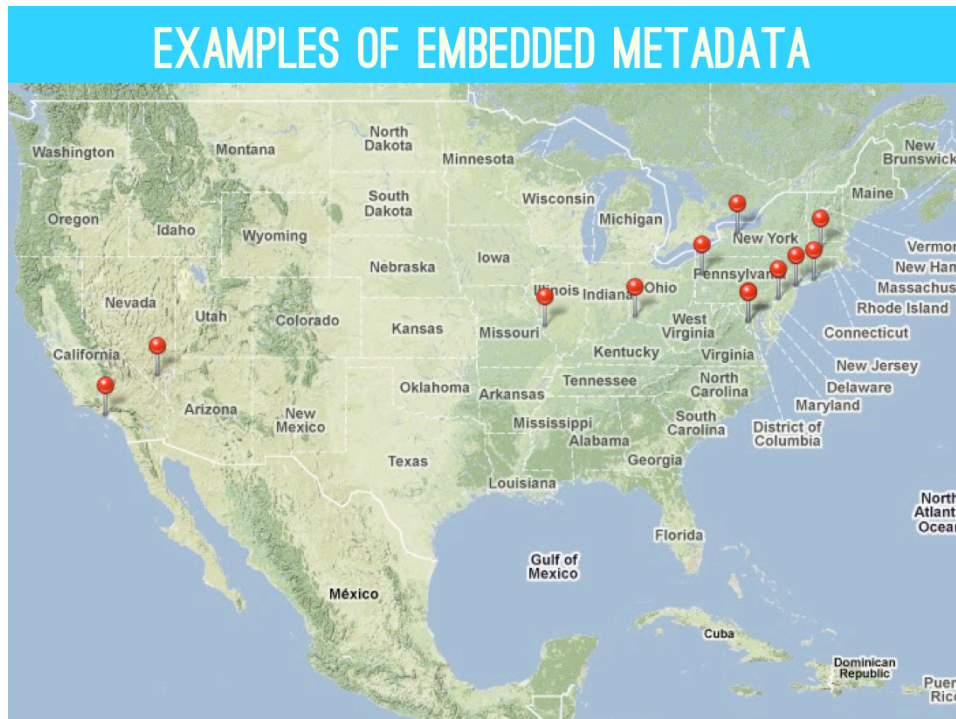
```
ExifTool Version Number      : 8.30
File Name                    : 5245486755_20acdc626_o.jpg
Directory                   : /Users/chris/Downloads
File Size                    : 3.7 MB
File Modification Date/Time  : 2011:05:31 00:45:44-04:00
File Permissions             : rw-r--r--
File Type                    : JPEG
MIME Type                    : image/jpeg
JFIF Version                 : 1.01
Exif Byte Order              : Little-endian (Intel, II)
Image Description            : President Barack Obama looks over the view of the ci
                             ty from the InterContinental Yokohama Grand Hotel in Yokohama, Japan, Nov. 13, 2010. (
                             Official White House Photo by Pete Souza)
Make                         : Canon
Camera Model Name            : Canon EOS 5D Mark II
X Resolution                 : 240
Y Resolution                 : 240
Resolution Unit              : inches
Modify Date                  : 2010:12:01 10:58:42
Artist                       : Pete Souza
Copyright                    : This photograph is provided by THE WHITE HOUSE as a
                             courtesy and may be printed by the subject(s) in the photograph for personal use only.
                             The photograph may not be manipulated in any way and may not otherwise be reproduced,
                             disseminated or broadcast, without the written permission of the White House Photo Of
                             fice. This photograph may not be used in any commercial or political materials, advert
                             isements, emails, products, promotions that in any way suggests approval or endorsemen
                             t of the President, the First Family, or the White House.
Exposure Time                : 1/200
F Number                     : 11.0
Exposure Program             : Manual
```

this, which is an ExifTool output.

EXAMPLES OF EMBEDDED METADATA

Name	Artist	Time	Date Added	Album
<input checked="" type="checkbox"/> File 3 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 3 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 3 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 4 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 4 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 4 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 5 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 5 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 5 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 6 copy 3		0:04	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 6 copy 4		0:04	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 6 copy 5		0:04	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 7 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 7 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 7 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 9 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 9 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 9 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File1 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File1 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File1 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File2 copy 3		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File2 copy 4		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File2 copy 5		0:05	5/20/10 11:01 PM	
<input checked="" type="checkbox"/> File 3		0:05	5/20/10 10:59 PM	
<input checked="" type="checkbox"/> File 3 copy		0:05	5/20/10 10:59 PM	
<input checked="" type="checkbox"/> File 3 copy 2		0:05	5/20/10 10:59 PM	
<input checked="" type="checkbox"/> File 4		0:05	5/20/10 10:59 PM	
<input checked="" type="checkbox"/> File 4 copy		0:05	5/20/10 10:59 PM	

This demonstrates what iTunes would like with bad/no embedded metadata in files.



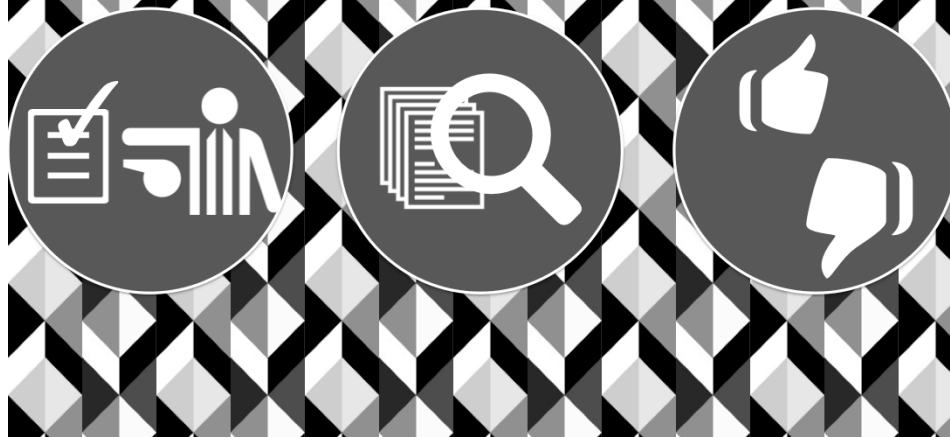
And applications using embedded metadata allow things like this where iPhoto takes embedded GPS data in photos and videos and plots them on a map.

WHAT DOES IT DO?



Analyzes a batch of files, using MedialInfo or Exiftool

WHAT DOES IT DO?



Reports successes and failures against identified ruleset



Because you have specifications and you need to make sure you keep an eye on them.

d. DESTINATION FILE FORMAT SPECIFICATIONS
Identify the target format specifications in the necessary detail to avoid miscommunication. For instance:

A Preservation Master, Access Master, and Access Copy shall be created for each tape, according to the specifications detailed below.

i. Preservation Master
 QuickTime wrapper (.mov extension)
 Video encoded using 10-bit YUV 4:2:2 uncompressed codec with the FourCC 'v210'.
 Audio encoded as uncompressed PCM
 Maintain the original aspect ratio, record channels and auxiliary information such as

ii. Access Master
 QuickTime wrapper (.mov extension)
 Video encoded using the FourCC 'dv' codec
 Audio encoded as uncompressed PCM
 Maintain the original aspect ratio, record channels and auxiliary information such as

iii. Access Copy
 QuickTime wrapper (.mov extension)
 Video encoded using the H.264 codec
 Audio encoded as uncompressed AAC

11. METADATA
Identify your metadata requirements. Specify the fields, vocabularies, definitions, controls and provide example values. Also specify the form of delivery and formatting required. For instance:

a. EXTERNAL
 CLIENT will furnish basic descriptive and technical metadata associated with the originals. All data shall be verified to be accurate and match the media throughout the reformatting process. Differences between furnished data and data and/or content on the originals shall be brought to the attention of CLIENT before dissemination of the data.
 The metadata fields and structure required for capture and delivery by the vendor are defined in the attached schema titled "CLIENT_reformatting_metadata_spec.xls"

b. EMBEDDED
 The following data shall be embedded into the designated field within the QuickTime container for each file. All information with the exception of Reference File will be provided by the client:

Information	QT field
Full Filename with extension	Description
Organization Name	Producer
"Reference File" Corresponding full reference filename	Comment
Source Item ID	Information
Rights Statement	Copyright
Title information	Title

Example of a statement of work. Both the file format specifications and the metadata specifications result in embedded metadata that you can/ should perform quality control on.



MedialInfo and Exiftool can be used but to really use these tools in the way we are here you need to be a power user of those tools and of other tools to string together the workflow in an automated way. So, many people have done this work manually or aren't doing it all because it's too resource intensive. Doing it in an automated way saves a tremendous amount of time. We had been developing complex scripts to help clients do QC, and they required a level of expertise to run and maintain that were unreasonable. This led to MDQC.



See [S12-Lacinak_mdqc_demo.mov](#)



Find these tools and others at www.avpreserve.com/tools

