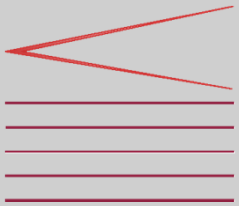

george blood audio

. . . and video

Describing Ourselves to Death:
The Failures of Metadata

- George Blood Audio, LP

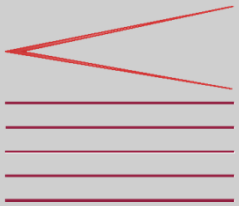
National Digital Stewardship Alliance
January 2014, Philadelphia

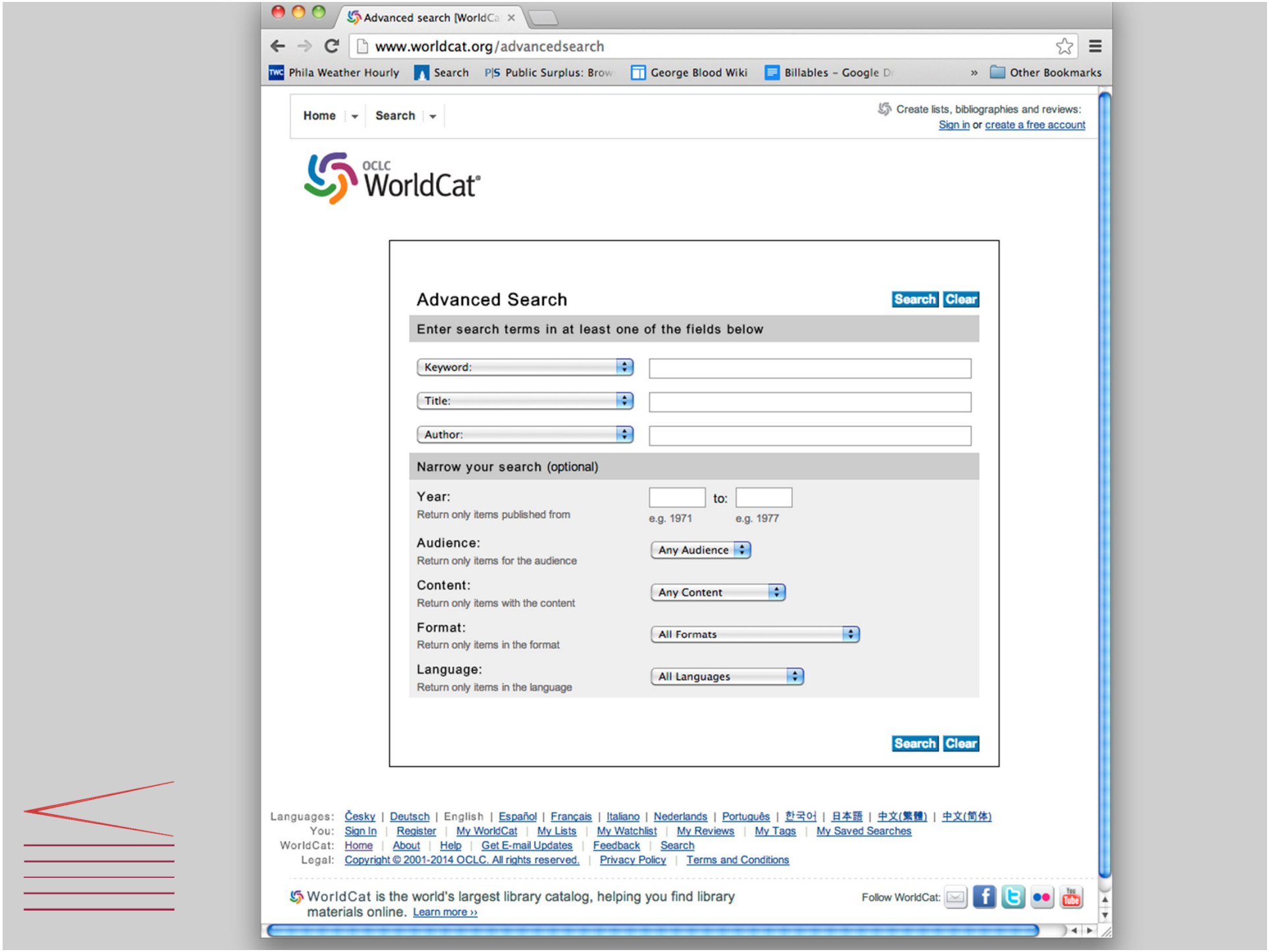


Definition by ALA PARS

Digital Preservation:

"Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the **accurate rendering of authenticated content over time**."

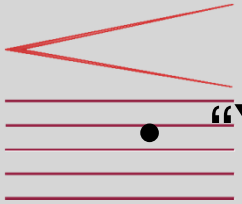



Advanced search [WorldCa
www.worldcat.org/advancedsearch
Phila Weather Hourly
Search
Public Surplus: Brow
George Blood Wiki
Billables - Google Dr
Other Bookmarks
Home
Search
Create lists, bibliographies and reviews:
Sign in or create a free account
OCLC WorldCat
Advanced Search
Search
Clear
Enter search terms in at least one of the fields below
Keyword:
Title:
Author:
Narrow your search (optional)
Year:
Return only items published from
to:
e.g. 1971
e.g. 1977
Audience:
Return only items for the audience
Any Audience
Content:
Return only items with the content
Any Content
Format:
Return only items in the format
All Formats
Language:
Return only items in the language
All Languages
Search
Clear
Languages: Česky | Deutsch | English | Español | Français | Italiano | Nederlands | Português | 한국어 | 日本語 | 中文(繁體) | 中文(简体)
You: Sign In | Register | My WorldCat | My Lists | My Watchlist | My Reviews | My Tags | My Saved Searches
WorldCat: Home | About | Help | Get E-mail Updates | Feedback | Search
Legal: Copyright © 2001-2014 OCLC. All rights reserved. | Privacy Policy | Terms and Conditions
WorldCat is the world's largest library catalog, helping you find library materials online. Learn more ››
Follow WorldCat:

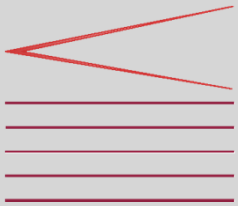

COMO HOPPER

In the words of Grace Hopper...

- “It's easier to ask forgiveness than it is to get permission”
- “A ship in a harbor is safe, but that is not what a ship is built for”
- “From then on, when anything went wrong with a computer, we said it had bugs in it”
- “You manage things; you lead people”



"The great thing about standards is that there are so many to choose from."



Standards are like toothbrushes.

Everyone agrees they're desirable…

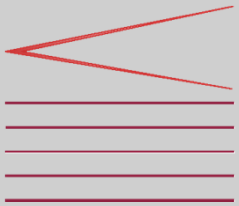
Standards are like toothbrushes.

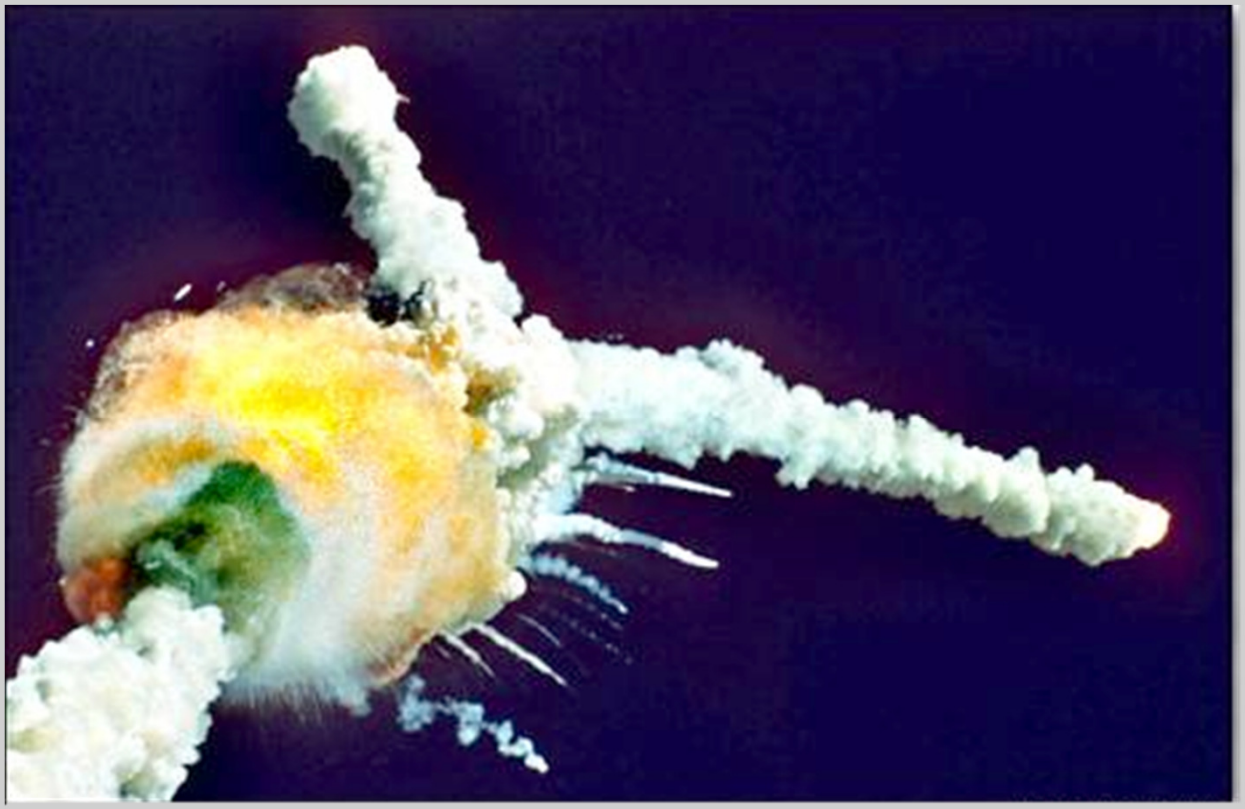
Everyone agrees they're desirable…

but nobody wants to use someone else's.

Why are we collecting all this metadata?

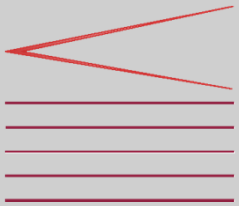
- To provide for discovery
- To manage the files
- To provide provenance
- To provide authenticity
- Etc.





Fundamental Problem

- Institutions don’t ask
  - “What problem are we protecting against”, or
  - “What function do we need to provide for”?

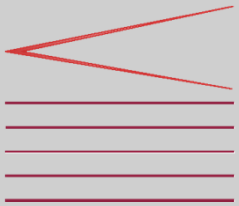



HORNET + 3


NCAP 35 mph TESTING
2003 CADILLAC CTS

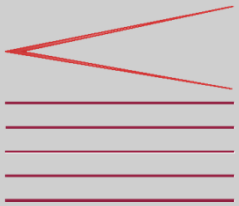
Metadata

- ≠ Cataloging and Description
  - Cataloging and Description aid discovery
  - Metadata allows the data to function in systems
- How much is enough?
- Is it possible to have too much?
- Why do we need more than we did before?
  - Are we moving the goal posts?
  - To what extent are our neuroses about digital preservation a reflection of our failures in analog preservation?
  - Is more metadata less product? By doing “better” for one object are we preserving less overall?
- Has anyone asked the users what they need?



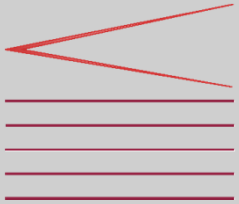
Organizing metadata

- “Standards”
- Toothbrushes

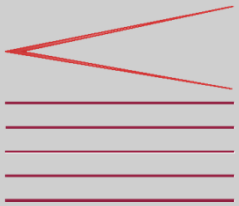


What is a standard?

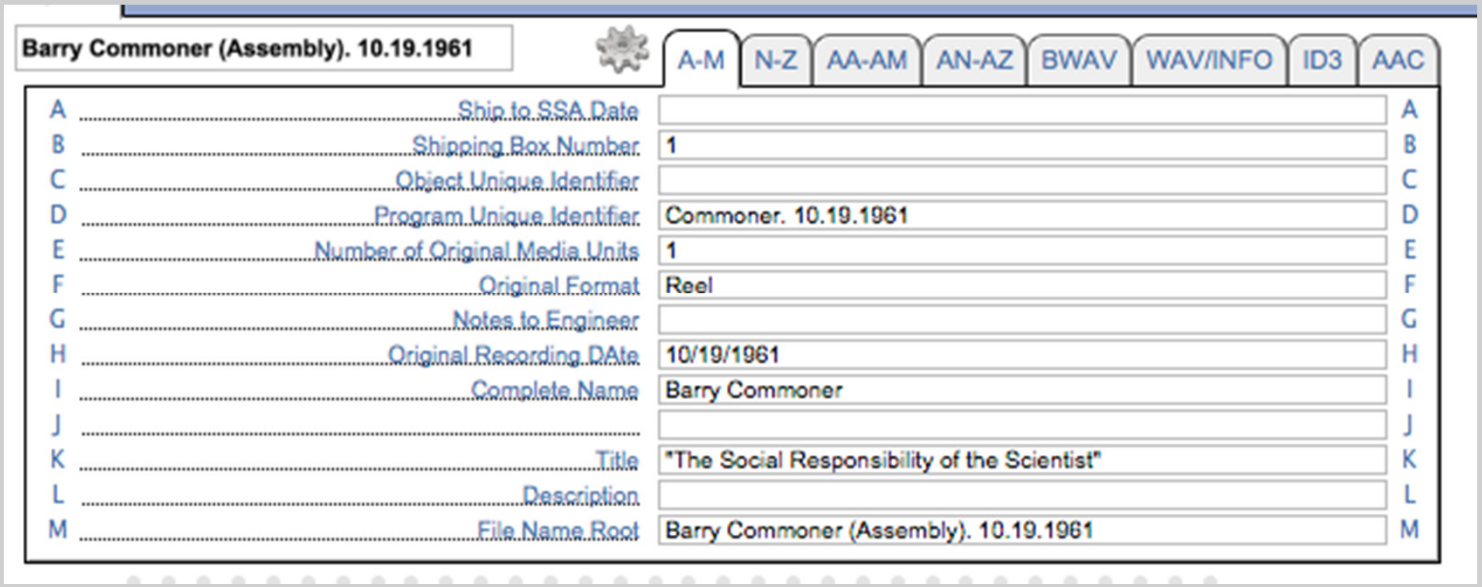
- How widely adopted?
- If everyone is doing something... is that good enough to be a “standard”?
- Does a standard have to be perfect?
- Does one size fit all?
- If there’s a standard and no one uses it, what’s it matter?
- What are the implications if there’s a standard and it is “locally modified”?
- If you make your own “standard”, in what ways does this enhance or inhibit preservation and long-term access?
  - **Aren’t we taught to avoid proprietary solutions? Why not for metadata?**



SIPS:
The State of the “Art”



Oberlin metadata

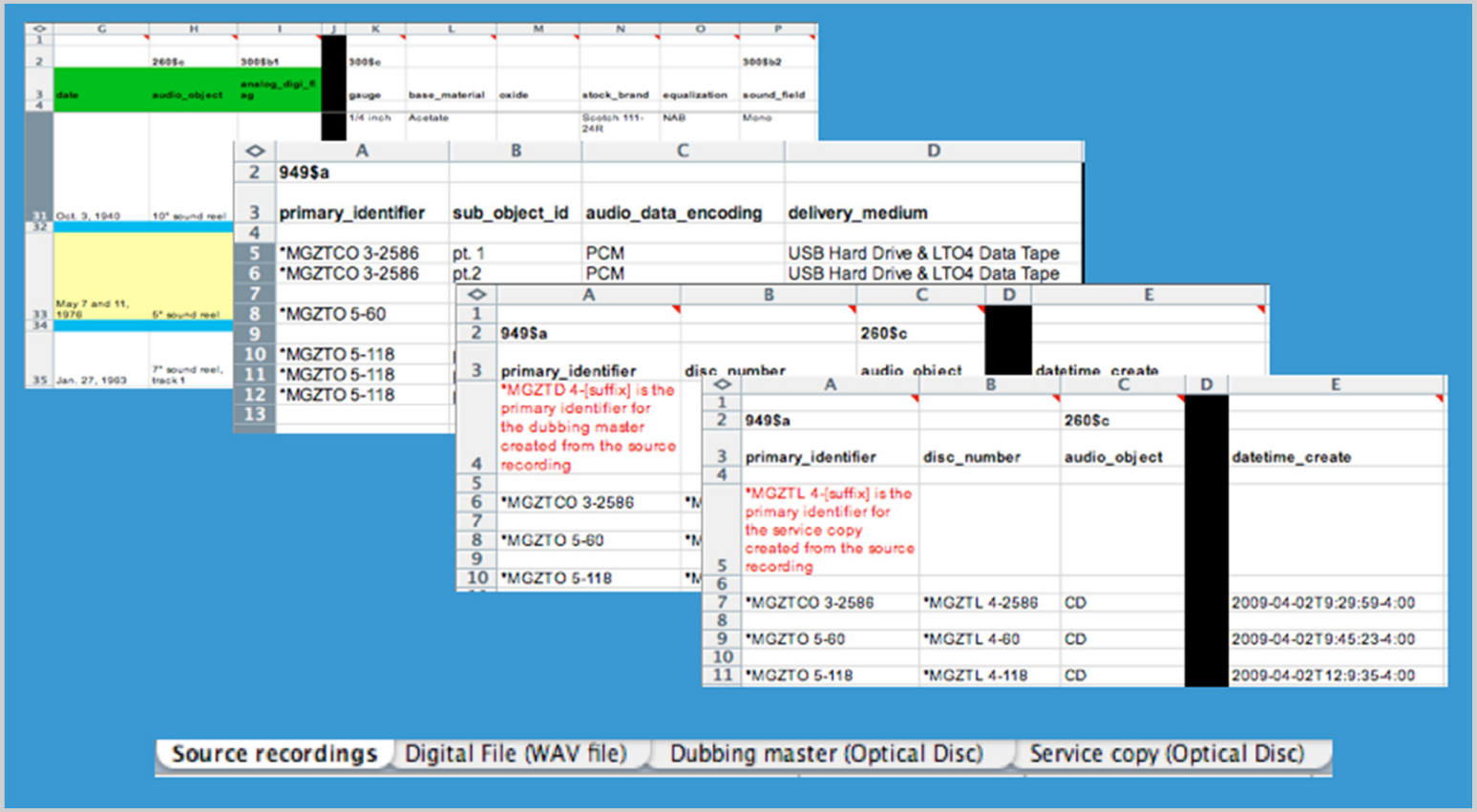


Barry Commoner (Assembly). 10.19.1961

A-M | N-Z | AA-AM | AN-AZ | BWAV | WAV/INFO | ID3 | AAC

A	Ship to SSA Date		A
B	Shipping Box Number	1	B
C	Object Unique Identifier		C
D	Program Unique Identifier	Commoner. 10.19.1961	D
E	Number of Original Media Units	1	E
F	Original Format	Reel	F
G	Notes to Engineer		G
H	Original Recording DAte	10/19/1961	H
I	Complete Name	Barry Commoner	I
J			J
K	Title	"The Social Responsibility of the Scientist"	K
L	Description		L
M	File Name Root	Barry Commoner (Assembly). 10.19.1961	M

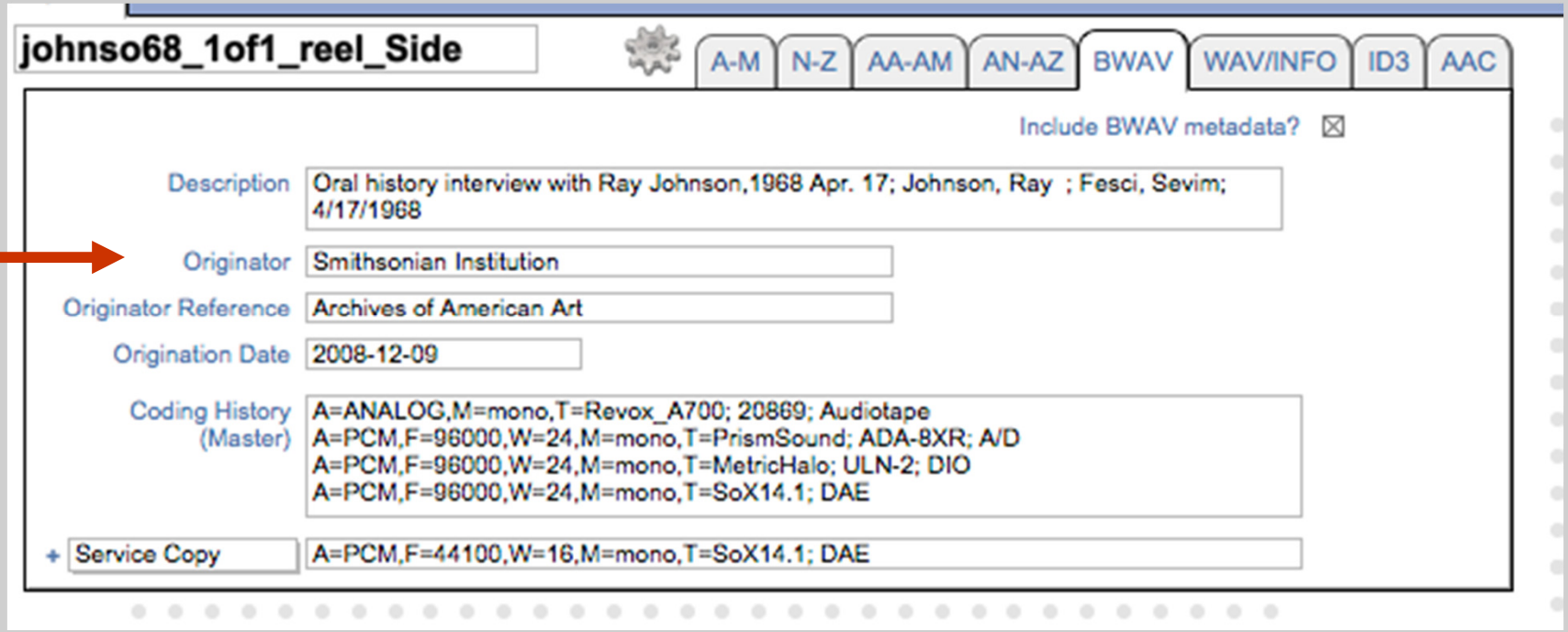
NYPL - LPA metadata



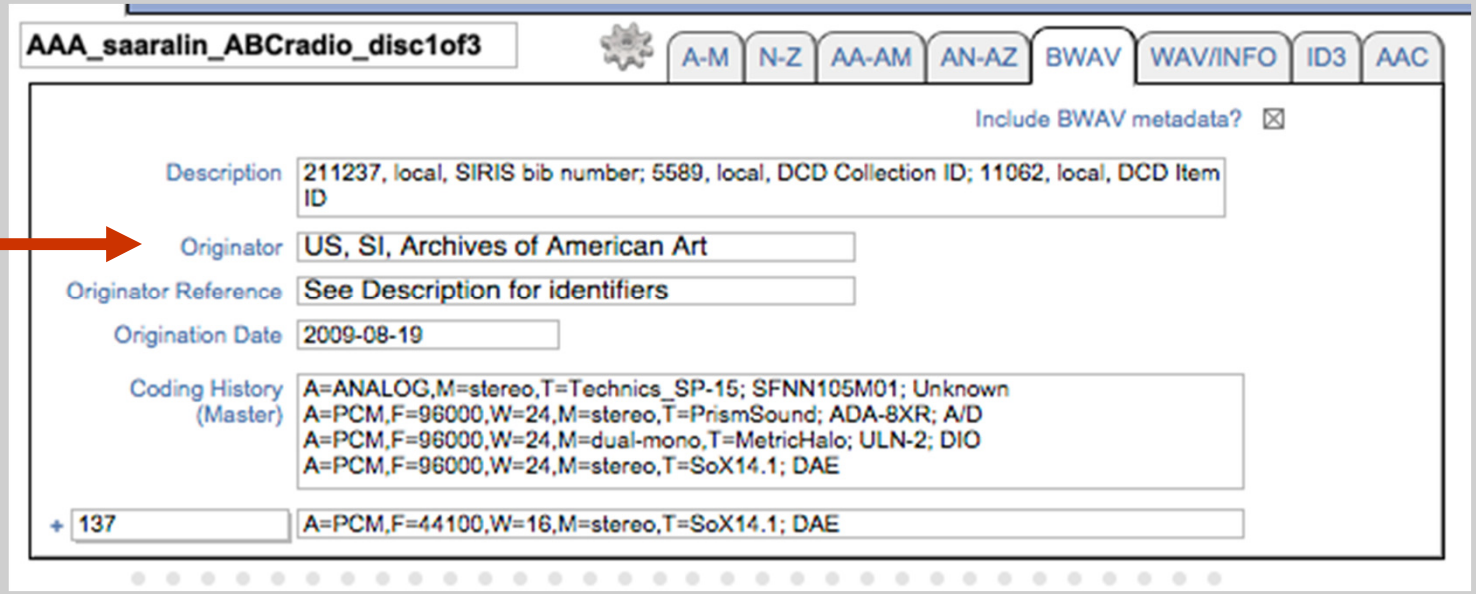
UMichigan RFI

University of Michigan Library	Audio Digitization Metadata List		Updated July 21, 2009				
Field	**Relation**	**Definition**	**Example**	**Required Status**	**Population**	**Origin**	
analog_digi_flag	Source recording	describes the method by which a physical audio object was recorded	Analog or Digital	Mandatory	U of M	LC	
dimensions_diameter	Source recording	audio object's diameter (in inches)	10 inches	Mandatory, if applicable	U of M	LC	
dimensions_height	Source recording	audio object's height (in inches)	4 inches	Mandatory, if applicable	U of M	LC	
dimensions_width	Source recording	audio object's width (in inches)	3 inches	Mandatory, if applicable	U of M	LC	
originating_library	Source recording	Library from U of M of which the source recording is a part.	SCL (Special Collections)	Mandatory	U of M	UM	
originating_collection	Source recording	Collection from U of M of which the source recording is a part	Rossiter, Wilson/Welles	Mandatory	U of M	UM	
generation	Source recording	describes the physical audio object	studio master, master, dub, original disc, etc.	Optional	U of M	LC	
audio_object	Source recording	an audio object's generic format name	LP, audio cassette, DAT, etc.	Mandatory	U of M (with vendor override)	HVD	
condition_note	Source recording	description of the state of a source recording's physical condition		Mandatory	U of M (with vendor override)	HVD	
audio_data_encoding	Digital file	structure for digital audio data	Pulse Code Modulated (PCM)	Mandatory	U of M	LC	
file_locat_value	Digital file	location of digital file within U of M	TBD	Mandatory	U of M	LC	
file_name	Digital file	Identifier of digital file	Barcode + face/track (39015234567890001)	Mandatory	U of M provides barcode / Vendor generates the latter	UM	
base_material	Source recording	a recording's base material	glass, aluminum, polycarbonate, unknown, etc.	Mandatory	Vendor	HVD	
dye_layer	Source recording	describes the dye present in recordable optical discs	phthalocyanine, cyanine	Mandatory if applicable	Vendor	NYPL	
equalization	Source recording	specific name of recording's inherent equalization (pre-emphasis)	NAB, Type I, Type II, unknown, etc.	Mandatory if applicable	Vendor	HVD	
gauge	Source recording	pertains to audio tape (expressed in inches)	1/4", 1/2", etc.	Mandatory if applicable	Vendor	HVD	
groove_orientation	Source recording	pertains to analog grooved media	Lateral or Vertical	Mandatory if applicable	Vendor	HVD	
sampling frequency	Digital file	rate at which audio was sampled for digital file	96K, 48K, 44.1K, etc.	Mandatory	Vendor	LC	
format_name	Digital file	official name of the file format	Broadcast Wave Format	Mandatory	Vendor		
note		any additional notes about the source recoding, the preservation master file, production master file or access copy	tracks, titles, timing, editing, processing	As necessary	Vendor	LC	

SI AAA Metadata



SI AAA Second Project



SI Hirshhorn and SI AAA

Sample Rate:
96000

Bit Depth:
24

Duration:
0:42:19

INFO Name:
Hess, Thomas B. "The Breakthrough of Abstract Expressionism."

INFO Artist:

INFO Date:
20090908

INFO Archival Location:
Smithsonian Institution Libraries, Hirshhorn Museum Library

INFO Copyright:
Material may be protected by copyright. Restrictions may apply.

BEXT Description:
Hess, Thomas B. "The Breakthrough of Abstract Expressionism."
Lecture at NGA, 11-4-73: 0001, File Identifier; HMSG0001A-B, Tape Identifier

BEXT Originator:
Hirshhorn Museum Library

BEXT Originator Reference:

BEXT Origination Date:
2009-09-08

BEXT Time Reference:
0

BEXT Version:
1

BEXT Coding History:
A=ANALOG,M=stereo,T=Nakamichi_Dragon; 09095; TDK_C90
A=PCM,F=96000,W=24,M=stereo,T=PrismSound; ADA-8XR; A/D
A=PCM,F=96000,W=24,M=dual-mono,T=MetricHalo; ULN-2; DIO
A=PCM,F=96000,W=24,M=stereo,T=SoX14.1; DAE

Sample Rate:
96000

Bit Depth:
24

Duration:
0:56:32

INFO Name:

INFO Artist:

INFO Date:

INFO Archival Location:

INFO Copyright:

BEXT Description:
Oral history interview with Tony Rosenthal, 1968 May 10-June 29.;
Tony; Sevim; 1968 May 10-June 29

BEXT Originator:
Smithsonian Institution

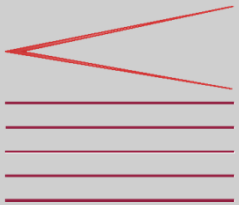
BEXT Originator Reference:
Archives of American Art

BEXT Origination Date:
2009-09-22

BEXT Time Reference:
0

BEXT Version:
1

BEXT Coding History:
A=ANALOG,M=mono,T=Revox_A700; 13652; Audiotape_1251
A=PCM,F=96000,W=24,M=mono,T=PrismSound; ADA-8XR; A/D
A=PCM,F=96000,W=24,M=mono,T=MetricHalo; ULN-2; DIO
A=PCM,F=96000,W=24,M=mono,T=SoX14.1; DAE

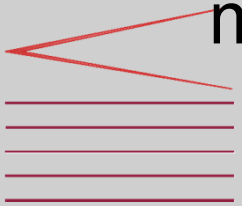


CUL METS


```
- <mets:mets OBJID="Goldstein_JonahJ_7139260_01" TYPE="audio_object_source">
  - <mets:metsHdr CREATEDATE="2009-10-15T15:22:19-04:00">
    - <mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
        <mets:name>Safe Sound Archive</mets:name>
      </mets:agent>
    - <mets:agent ROLE="PRESERVATION" TYPE="ORGANIZATION">
        <mets:name>Columbia University Libraries</mets:name>
      </mets:agent>
    </mets:metsHdr>
  - <mets:dmdSec ID="dmd1">
    - <mets:mdWrap MDTYPE="DC">
      - <mets:xmlData>
        - <oai_dc:dc>
          - <dc:title>
              Jonah J. Goldstein Oral History Interview : Source Audio Recording, no. 1
            </dc:title>
            <dc:identifier>Goldstein_JonahJ_7139260_01</dc:identifier>
            <dc:type>sound</dc:type>
          </oai_dc:dc>
        </mets:xmlData>
      </mets:mdWrap>
    </mets:dmdSec>
  - <mets:amdSec ID="amd1">
    - <mets:techMD ID="techmd1">
      - <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="JhoveXMLOutput">
        - <mets:xmlData>
          - <jhove xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/jhove http://hul.harvard.edu/ois/xml/xsd/jhove/1.5/jhove.xsd"
            name="Jhove" release="1.1" date="2008-02-21">
              <date>2009-10-15T15:22:19-04:00</date>
            - <repInfo uri="/Volumes/GBAUDIO/SSA/2007511%20CUL/Production/Goldstein_JonahJ_7139260/Master
              /Goldstein_JonahJ_7139260_01_01_m.wav">
                <reportingModule release="1.3" date="2007-12-14">WAVE-hul</reportingModule>
                <lastModified>2009-08-12T14:45:39-04:00</lastModified>
                <size>2057587049</size>
                <format>WAVE</format>
                <status>Well-Formed and valid</status>
              - <sigMatch>
                  <module>WAVE-hul</module>
                </sigMatch>
              - <messages>
                  <message offset="848" severity="info">Chunk type 'PAD ' ignored</message>
                </messages>
                <mimeType>audio/x-wave</mimeType>
```

How will any of this provide for discovery, management, provenance, etc?

- It all has to be done manually.
- It is just as much work to create software tools to read the metadata as to make it.
- It costs more to do the metadata work on some projects than the digitization.
- What will be the cost to reformat the metadata when the digital file is migrated?



Open Source!

Open Standards!!

Interoperability!!!

Except MY Metadata

DIPs: Let’s get religion