Oral History, METS and Fedora:

Building a Standards-Compliant Audio Preservation Infrastructure

COLUMBIA UNIVERSITY LIBRARIES / INFORMATION SERVICES

Outcomes of Columbia University Libraries' 2008-2010 Mellon-funded Audio Preservation Project

- Stephen Davis
  - Director, Libraries Digital Program Division
- Janet Gertz
  - Director, Preservation & Digital Conversion Division

What we'll cover

- Janet
  - Project background
  - Identifying and describing content and versions
  - Physical organization vs. intellectual organization
- Stephen
  - Metadata
  - Digital asset management, preservation, access
  - Conclusions

The Problem

Mellon-funded survey

in 2005-2007 found 35,000 pieces of unique analog audio aging rapidly in Columbia special collections

http://library.columbia.edu/services/preservation/audiosurvey.html

side 1

APC
90
SIDE
1
Professional
Mylar tape
Low noise, E
Royal cassette
Royal Sound Co., Inc.. Freeport, N.Y. 11520 U.S.A.

Audio preservation standards as of 2008

- Standards for digitization of audio well established
  - 96 kHz, 24 bit, Broadcast Wave format
- Standards for structural, technical, and preservation metadata still evolving
  - No clear model for METS for digitized audio
  - Audio Engineering Society draft standard

AES-X098
SC-03-06-B

DRAFT
Standard for audio preservation and restoration - Administrative and structural metadata for audio objects

Abstract

This standard provides a vocabulary to be used in describing structural and administrative metadata for digital and analog audio formats for the purpose of enabling audio preservation activities on those objects. Some implementations also refer to this metadata as technical metadata. The characteristics of the audio objects captured by this standard may be of use to audio communities beyond the audio preservation community.

Close Full Screen

AES-X098B Superseded By:

- AES57-2011-f (2011)
  AES standard for audio metadata - Audio object structures for preservation and restoration
- AES60-2011-f (2011)
  AES standard for audio metadata - Core audio metadata

Sound Directions: Best Practices for Audio Preservation

This publication presents the results of research and development carried out by the *Sound Directions* project with funding from the National Endowment for the Humanities in the U.S. Each chapter in this document is divided into two major parts: a preservation overview that summarizes key concepts for collection managers and curators, followed by a section intended for audio engineers, digital librarians, and other technical staff that presents recommended technical practices while summarizing our findings and experience.

Publication Download (PDF format, 5.52MB)

The appendices to the publication provide additional technical detail for those interested. There are five appendices:

- Appendix 1: Metadata Elements in the Audio Technical Metadata Collector (ATMC) (522KB)
- Appendix 2: XML Produced by Harvard's Audio Object Manager for Digital File Metadata (XML format - 6KB)
- Appendix 3: Sample Harvard METS Document (XML format - 565KB)
- Appendix 4: Sample Indiana METS Document (XML format - 133KB)
- Appendix 5: Software Tools in the Harvard *Sound Directions* Toolkit (96KB)

NOTE: Appendices two through four are examples of metadata documents in their raw XML form. Some web browsers will present this information in a separate window as text, revealing the structure of the document immediately. Some browsers will not and must be prompted to do so. In this case, please ensure that your web browser is set to "view source." You may also prefer to download the raw XML files and view them using XML editing software.

Contact Us

» soundir@indiana.edu

Purpose of Columbia's project

- Build a sustainable program for audio preservation at Columbia
- Reduce need for time-consuming custom metadata and ingest work by Libraries Digital Program Division staff
- Improve efficiency and consistency

Programmatic goals

- Aim for same quality product as achieved by Harvard & Indiana in *Sound Directions*
- **But** — digitization and metadata creation by external vendors
- Establish CUL infrastructure
  - Quality control procedures
  - Metadata requirements
  - Ingest into Fedora

Preservation priorities

Risk of loss	Intellectual value: High	Intellectual value: Low
High		
Low		

Columbia Center for Oral History

Archive

At the center of our work is our "living archive" of more than 8,000 aural and visual interviews that explore diverse topics in United States and global history.

News

Follow our blog for the most up-to-date reporting on CCOH's projects, collections, and events.

The Tool of History: A Public Discussion Tuesday, 12/6, at 6PM

50 Years: Amnesty International & Human Rights Advocacy Conference, Dec. 8 at 3pm

[Video] "United in Anger:

Research

CCOH conducts innovative oral history research projects, which we add to our archive and use as the basis for educational programs.

THE RULE OF LAW ORAL HISTORY PROJECT

Check out our newest oral history project, featuring digitally available transcripts of interviews on the topics of the death penalty and Guantanamo Bay.

9/11 ORAL HISTORY PROJECT

This September will mark the official opening of our September 11, 2001 Oral History Projects.

Education

CCOH is dedicated to building the field of oral history, and making our archives and expertise of use to the public.

NEW! Consult our Oral History Bibliography !

CCOH just wrapped up its annual

Oral History collections

- Our highest preservation priority
- Unique recordings held only at Columbia
- More than 8,000 interviews since 1948
- Over 15,000 physical objects
- Strong demand for access to the sound
- Many in poor condition

Project team

- Preservation & Digital Conversion Division
  - Project management, digitization, quality control
- Libraries Digital Program Division
  - R&D, METS, other metadata, Fedora ingest
- Columbia Center for Oral History
  - Selection, preparation, physical handling
- Bibliographic Control Division
  - Descriptive metadata, MARC records

george blood, L.P.

Audio/Video/Film Preservation

George Blood Audio
George Blood Video

—preserving the sound
and motion of history®

Preservation results

- 555 interviews preserved
  - 1,346 original audio objects
  - 2,100+ hours of sound
- 1,841 digital audio files created
- 555 MARC records created
- 555 sets of METS records created

Infrastructure results

- System for METS records to describe
  - Files that represent the original objects
  - Files that represent the intellectual objects
- Incorporation of draft Audio Engineering Society metadata into METS records
- Procedures for ingest into Fedora

Project challenges

- Describing versions and formats
- Identifying the content
- Coping with the disconnect between physical organization and intellectual organization

Oral histories are complicated

- Original audio recording on a series of tapes or cassettes
- Digitized audio: arranged to put all parts in chronological sequence
- Transcript: edited to suit the interviewee; doesn’t perfectly match the original audio
- Digitized transcript

The Reminiscences of

ADOLF A. BERLE, JR.

The Oral History Research Office

Columbia University

July 1974

ON SERVICES

assumed responsibility for the debts of the states and also for the creation of currency. I may add that that had a good deal of influence on my life later, because when -- in 1933, ~~when~~ we had to meet economic problems, I naturally thought of the Hamiltonian technique and tried to undergird the debt structure of the United States then with federal credit, ~~and~~ I sometimes wonder whether it derived from that original study with Ned Channing.

Q: Who were some of your classmates at the time? What I'd like to do is trace three strands here, one, the impact of your father on social legislation, the impact of teachers, and the group of students that you associated with at the time.

Berle: A man who is four years younger than the college generation has some difficulties. You lose as well as win. If you win time, you may lose in associations. Actually the relations I had, especially in my first couple of years, were with the seniors when I came in. They were too far ahead to be worried about the youngster,

COLUMBIA UNIVERSITY LIBRARIES / INFORMATION SERVICES

MARC records

- One record for analog versions
  - Paper transcript
  - Audio tapes, cassettes, etc.
- One record for digital versions
  - Digitized or born digital audio
  - Transcript in Word or other format

6 Oral history interview with Kitty Carlisle Hart, 1979.
Carlisle, Kitty, 1910-2007,

Call Number: NXCP88-A1159
Location: Oral History, 801 Butler (Non-Circulating)
No status information

7 Oral history interview with Kitty Carlisle Hart [electronic resource], 1979.
Carlisle, Kitty, 1910-2007,

Call Number: Carlisle, K.
Location: Oral History, 801 Butler (Non-Circulating)
No status information

8 Pre-code Hollywood collection

Oral history interview with Kitty Carlisle Hart,

Author: Carlisle, Kitty, 1910-2007, interviewee.
Title: Oral history interview with Kitty Carlisle Hart, 1979.
Description: Transcript: 805 leaves.
Tape: 1 reel, 12 cassettes.
Restrictions: Permission required for access.
Permission required to cite, quote, and reproduce. Contact repository for information.
LC Subjects: Carlisle, Kitty, 1910-2007--Interviews.
Hart, Moss, 1904-1961.
Singers--Interviews.
Actresses--Interviews.
Motion pictures and music.
Musicals.
Music--Instruction and study.
Theater.
Television.
New York (State)--Politics and government--1951-
New York (State)--Social life and customs.

Oral history interview with Kitty Carlisle Hart

Author: Carlisle, Kitty, 1910-2007, interviewee.
Title: Oral history interview with Kitty Carlisle Hart [electronic resource], 1979.
Description: 13 sound files : digital preservation master, WAV files (96 kHz, 24 bit)
Restrictions: Permission required for access.
Permission required to cite, quote, and reproduce. Contact repository for information.
LC Subjects: Carlisle, Kitty, 1910-2007—Interviews.
Hart, Moss, 1904-1961.
Singers—Interviews.
Actresses—Interviews.
Motion pictures and music.
Musicals.
Music—Instruction and study.
Theater.
Television.
New York (State)—Politics and government—1951-
New York (State)—Social life and customs.
Also Listed Under: Diamonstein-Spielvogel, Barbaralee, interviewer.

Holdings Information:

Identifying the content

All we know
is what someone
has written on the container

Reel No. _____

MAGNA-REEL also available in

- Lifetime, moisture-proof metal cans
- All metal storage chests (holds 12 tapes)
- 5", 7", 10" empty tape cans fitting other brands

Manufacturers: **RECORDING TAPE CO.** New York City 28—Newtown, Conn.

☑ Single Track ☐ Dual Track ☐ Binaural

LECT II

Track	Speed	Subject	Date	Footage	Recording Time
	7½	Reel 2	1/16/61		
		[illegible]			
		Done = 2/23/61 (Abraham Davidson)			

Other information sources

- Transcripts
- Card file of interviewees
- Paper files of correspondence with interviewees and interviewers
- Staff memories
- Listening to the audio after digitization

If content identification is inaccurate, projects are difficult and more expensive

- Quality control must be slow and 100%
- Metadata requires significant revisions
- Vendor has to make many changes
- Version control is essential

Disconnect between
physical organization
and
intellectual organization

Oral histories are complicated

Session: basic unit of an oral history

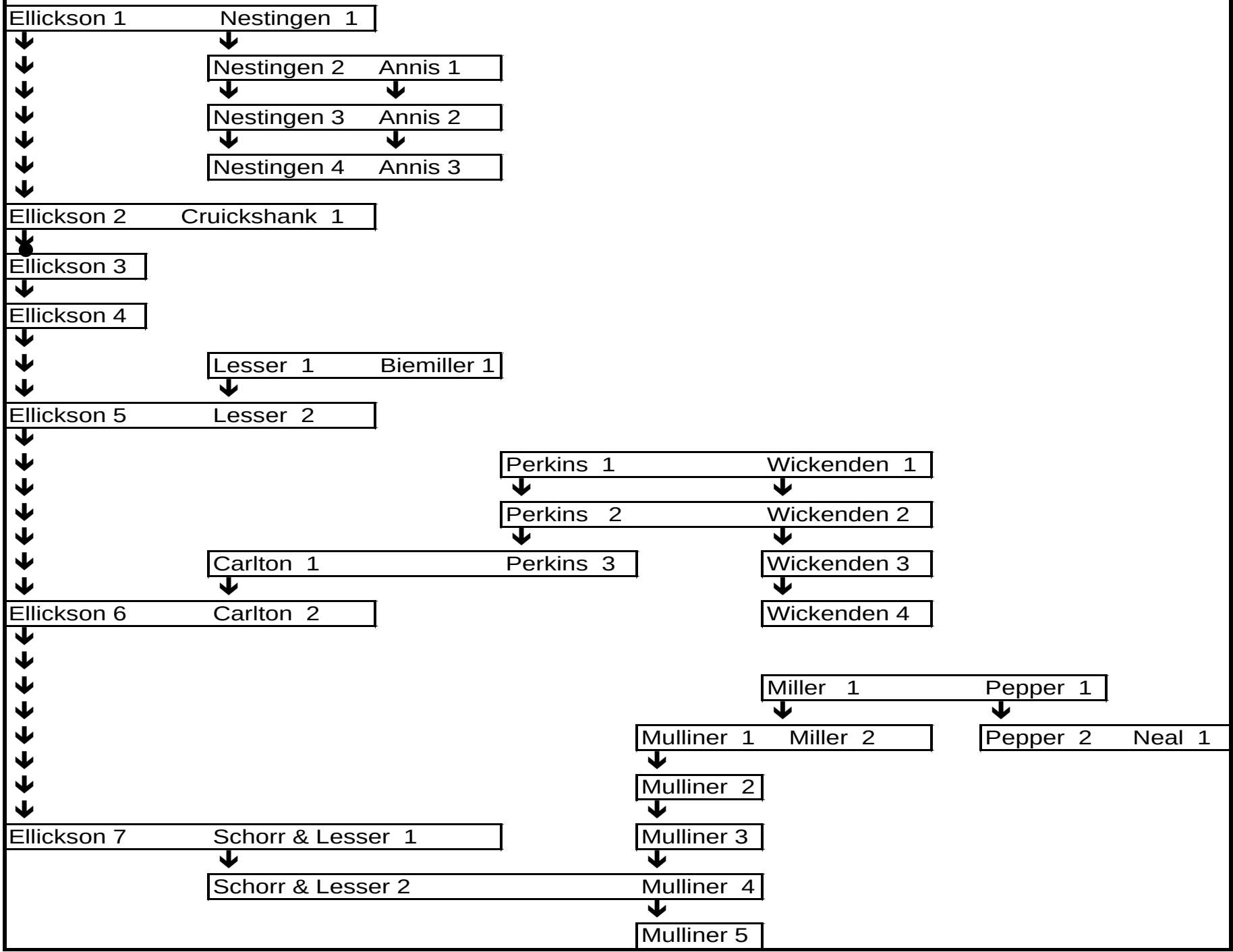
- Single recorded sitting of ca. 1.5-2 hours
- Can number a few or more than 20 in one oral history
- Recorded over a period ranging from days to years

Mind / body disconnect

- One session = one tape
- One session = several tapes

- One tape = one session
- One tape = several sessions from one oral history
- One tape = sessions from several oral histories

- Several tapes = several sessions from several oral histories


Ellickson 1
Nestingen 1
Nestingen 2
Annis 1
Nestingen 3
Annis 2
Nestingen 4
Annis 3
Ellickson 2
Cruickshank 1
Ellickson 3
Ellickson 4
Lesser 1
Biemiller 1
Ellickson 5
Lesser 2
Perkins 1
Wickenden 1
Perkins 2
Wickenden 2
Carlton 1
Perkins 3
Wickenden 3
Ellickson 6
Carlton 2
Wickenden 4
Miller 1
Pepper 1
Mulliner 1
Miller 2
Pepper 2
Neal 1
Mulliner 2
Ellickson 7
Schorr & Lesser 1
Mulliner 3
Schorr & Lesser 2
Mulliner 4
Mulliner 5

- Preservation practice: make a copy that accurately represents the original object
- Patron needs: a coherent sequence of files that contain all and only one oral history

Solution

- Preservation master file
  - Accurately captured from the physical object
  - 96 kHz, 24 bit, Broadcast Wave format
- Rendered file
  - Concatenates all parts of an interview regardless of which master files they're on
  - 96 kHz, 24 bit, Broadcast Wave format
- ADL (Audio Decision List)
  - Metadata that tracks which minutes from which master files make up the rendered file

Over to Stephen ….

Metadata

- MARC, AES, MODS, DUBLIN CORE, PREMIS
- ADL (AES *Audio Decision List*)
- METS (*Metadata Encoding and Transmission Standard*)
- RDF *(Resource Description Framework)*
- ORE (OAI *Object Reuse and Exchange*)

ADL (*Audio Decision List)*

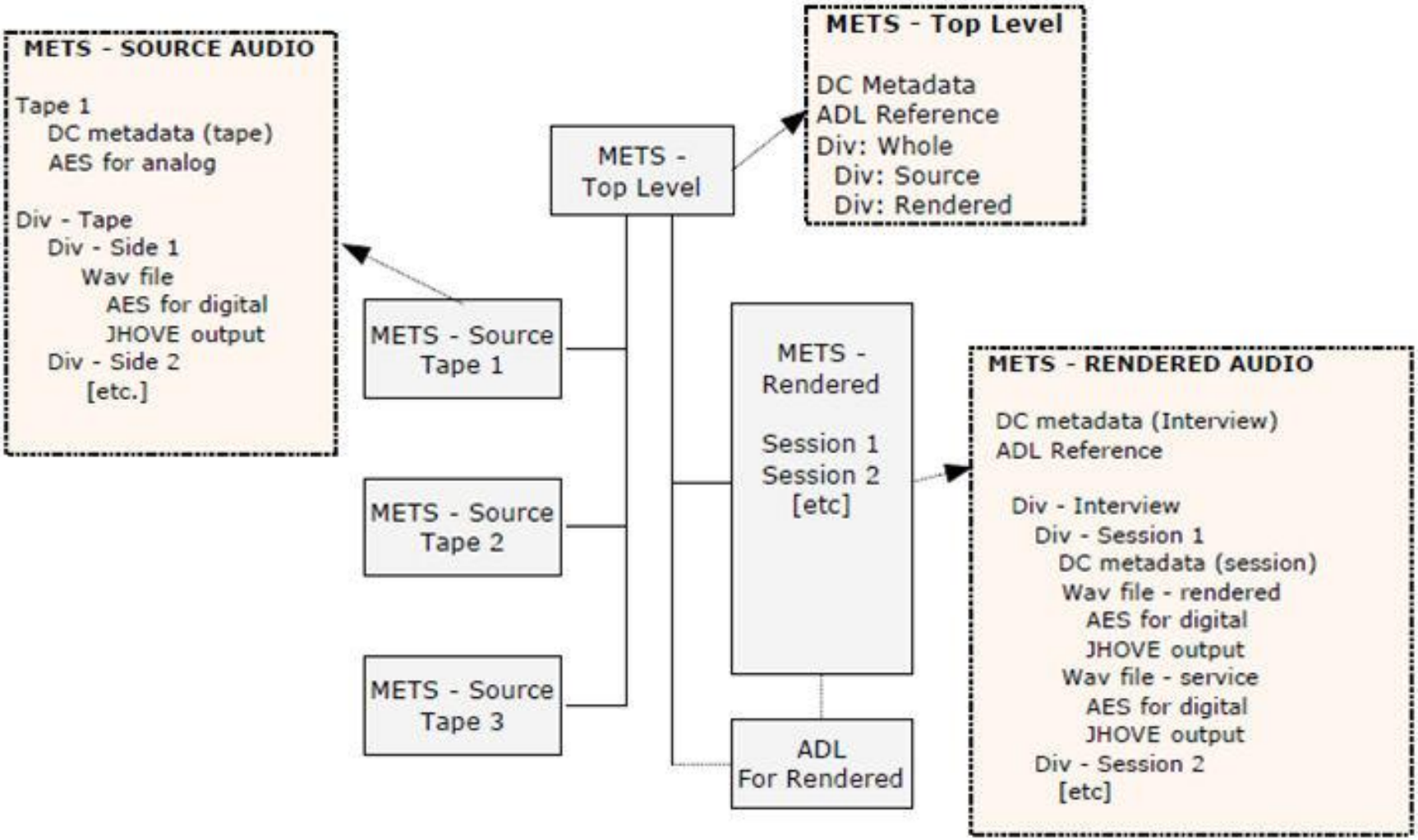
- ADL specified in AES 31-3 standard
- Records edit decisions
- Designed to be imported into audio editing software to recreate those decisions
- Imperfectly supported by commercial software platforms
- Migratable, human readable (sort of)

```
/* Generated by: WaveLab 6.10 (build 340) */
<ADL>
    <VERSION>
        (ADL_ID)    "06,64,43,52,01,01,01,04,01,02,03,04,"
        (ADL_UID)   A968ED9C-58A7-716B-ACE3-F404C25D1F30
        (VER_ADL_VERSION)   01.01.00.00.03
        (VER_CREATOR)   "WaveLab"
        (VER_CRTR)  6.10_(build_340)
    </VERSION>
    <PROJECT>
        (PROJ_TITLE)    "Anderson_Judith_6880292"
        (PROJ_ORIGINATOR)   "Steinberg Media Technologies"
        (PROJ_CREATE_DATE)  2009-03-11T15:23:13
        (PROJ_NOTES)    "_"
        (PROJ_CLIENT_DATA)  "_"
    </PROJECT>
    <SYSTEM>
        (SYS_SRC_OFFSET)    00.00.00.00*0000
        (SYS_BIT_DEPTH) 32
        (SYS_GAIN)  0.0
    </SYSTEM>
    <SEQUENCE>
        (SEQ_SAMPLE_RATE)   S96000
        (SEQ_FRAME_RATE)    25
        (SEQ_ADL_LEVEL) 1
        (SEQ_CLEAN) FALSE
        (SEQ_SORT)  0
        (SEQ_MULTICHAN) TRUE
        (SEQ_DEST_START)    00.00.00.00*0000
    </SEQUENCE>
    <TRACKLIST>
        (Track) 1   ""
        (Track) 2   ""
    </TRACKLIST>
    <SOURCE_INDEX>
        (Index) 1
            (F) "url:file://localhost/Y:/SSA/2007511 CUL/Production/Anderson_Judith_6880292/Master/Anderson_Judith_6880292_01_01_m.wav"
            "CLIO-6880292"  00.00.00.00*0000    00.47.12.01*0255    "_" "N"
    </SOURCE_INDEX>
    <EVENT_LIST>
        (Entry) 1
            (Cut)   I   1   1-2 1-2
                00.00.00.00*0000    00.00.01.04*3328    00.47.13.05*3583    R
            (Rem)   NAME    "Anderson_Judith_6880292_01_r"
    </EVENT_LIST>
    <WAVELAB_CHANNEL_MAPPING>
        mode k_dvda_stereo
        1L Lf
        2R Rf
    </WAVELAB_CHANNEL_MAPPING>
</ADL>
```

METS

- Relate master and rendered files to each other and to the ADL
- Include AES draft metadata for
  - Technical details of physical object
  - Technical details of digital object
- Technical details of capture process

Mellon Audio Preservation -- Metadata Overview

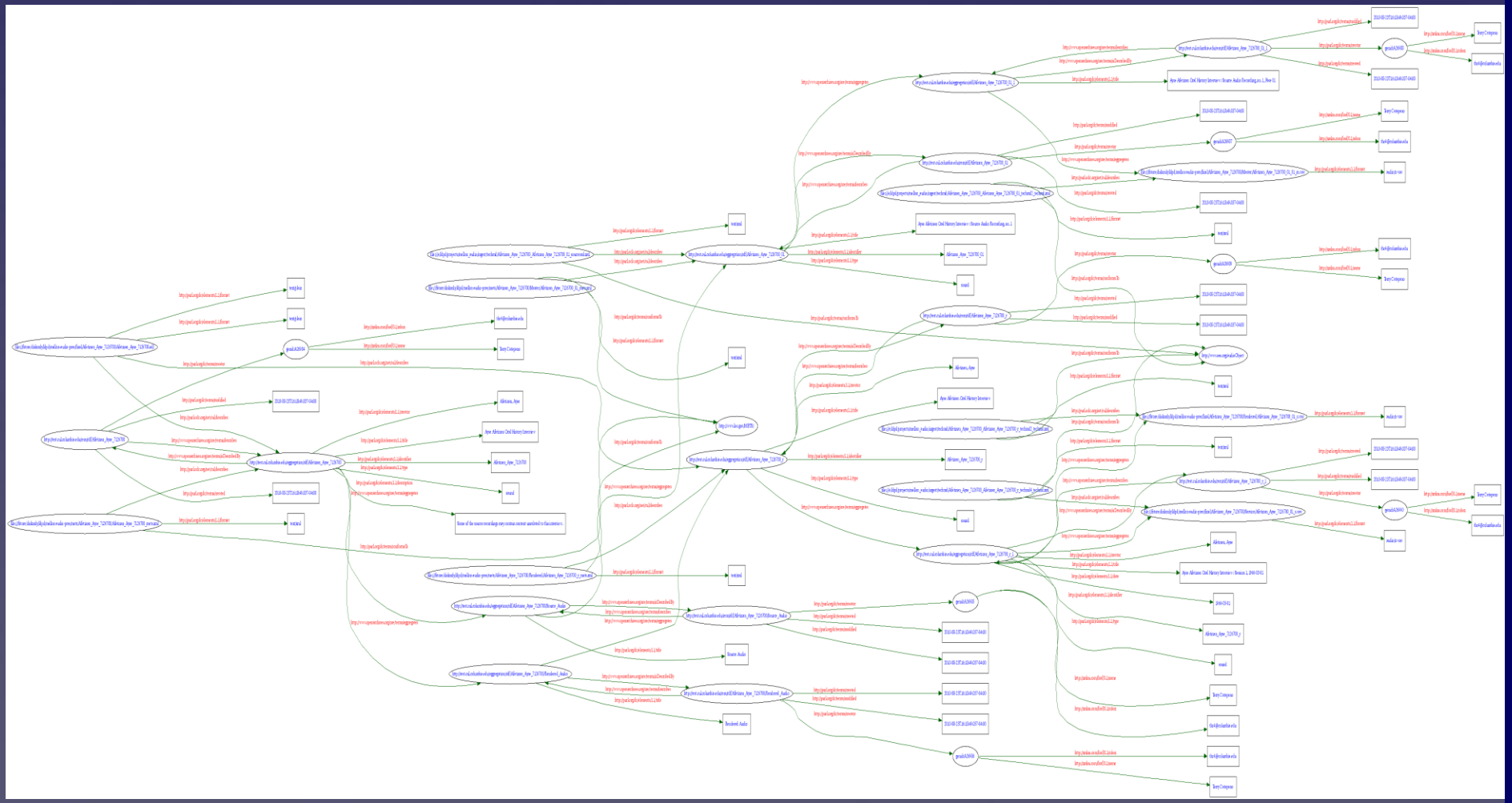


RDF (Resources Description Framework) Version of METS for Complete Interview

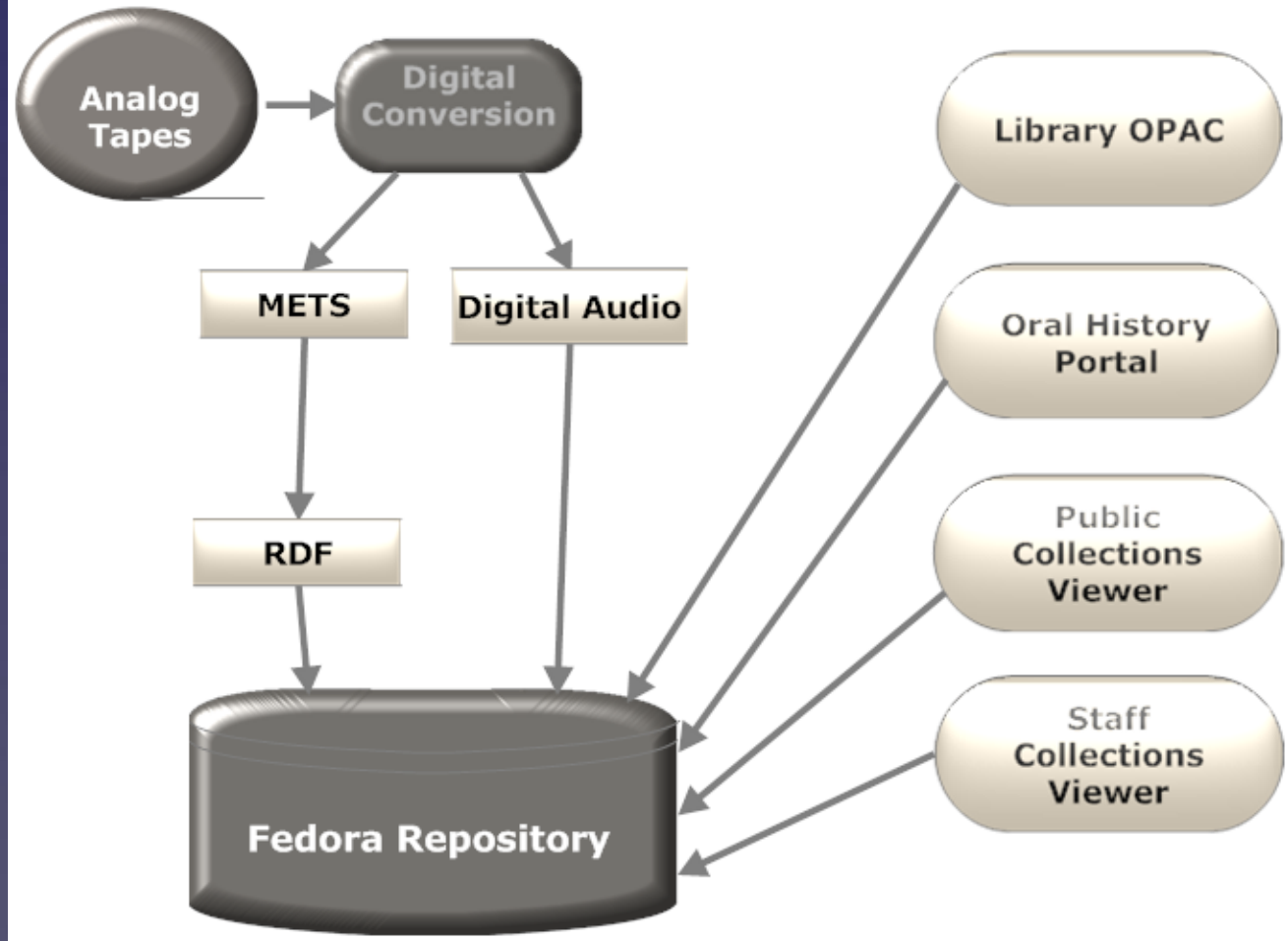
```
<?xml version="1.0" encoding="UTF-8"?>
-<rdf:RDF xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:mets="http://www.loc.gov/METS/"
xmlns:xd="http://www.oxygenxml.com/ns/doc/xsl"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:cul="http://purl.oclc.org/net/cul/"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:ore="http://www.openarchives.org/ore/terms/"
xmlns:dcterms="http://purl.org/dc/terms/">

<rdf:RDF><rdf:Description rdf:about="
http://test.cul.columbia.edu/rem/rdf/Trilling_Diana_7606158">
        <ore:describes rdf:resource="
        http://test.cul.columbia.edu/aggregation/rdf/Trilling_Diana_7606158"/>
        <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
        2011-10-29T13:55:46.564-04:00</dcterms:created>
        <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
        011-10-29T13:55:46.564-04:00</dcterms:modified>
        </rdf:Description>
<rdf:Description rdf:about="
http://test.cul.columbia.edu/aggregation/rdf/Trilling_Diana_7606158">
        <ore:isDescribedBy rdf:resource="
        http://test.cul.columbia.edu/rem/rdf/Trilling_Diana_7606158"/>
        <dc:creator>Trilling, Diana</dc:creator>
        <dc:title>Diana Trilling Oral History Interview</dc:title>
        <dc:relation>clio:7606158</dc:relation>
        <dc:type>sound</dc:type>
        <dc:description>Some of the source recordings may contain content unrelated to this
        interview.</dc:description>
        <ore:aggregates rdf:resource="
        http://test.cul.columbia.edu/aggregation/rdf/Trilling_Diana_7606158/Source_Audio"/>
        <ore:aggregates rdf:resource="
        http://test.cul.columbia.edu/aggregation/rdf/Trilling_Diana_7606158/Rendered_Audio"/>
</rdf:Description>
<rdf:Description rdf:about="
file:///fstore/archive/ldpd/preservation/mellon_audio_2010/data/Trilling_Diana_7606158/Trillin
g_Diana_7606158.adl">
        <dc:format>text/plain</dc:format>
        <cul:describes rdf:resource="
```

RDF Graph of Single Set of Interviews (Visual)



Oral History Audio Data Flow



Oral History Interviews Originally Targeted for Project *[beginning of list]*

Original Interviews Targeted for Mellon Audio Preservation Project			Jul-08			
OHROLastName	OHROFirstName	CollectionName	in CLIO	Notes	Quantity	MaxTime Minutes
Abram	Morris	Ethnic Groups and American Foreign Policy		3 sound tape reels.	3	143
Abzug	Bella	Biographical Interviews	no	17 sound cassettes.	17	1000
Aldewereld	Siem	World Bank		Sound tape reel.	1	240
Aldrich	Alice	Allan Nevins Project	no	1 Sound tape reel. With: Lillian Bean, Day Tuttle	1	120
Alexensky		Biographical Interviews	no	2 sound tape reels.	2	240
Alter	Gerald	World Bank		Sound tape reel. With John deWilde, Harold Lar	1	60
Amitay	Morris	Ethnic Groups and American Foreign Policy		5 sound tape reels. Notes: 5 Okay, 3 Missing.	5	151
Anderson	Judith	Hollywood Film Industry Project		1 Sound cassette.	1	90
Anderson	Quentin	Diana Trilling Project	no	3 sound tape reels. With: Stephen Donadio, The	3	387
Anderson	Thelma	Diana Trilling Project	no	3 sound tape reels. With: Midge Podhoretz, Frit	3	395
Andrews	U.J.	Black Journalists Project	no	1 sound tape reel. With: George McElroy.	1	180
Annis	Edward	Social Security		5 Sound tape reel. With : Ivan Nestingen, Caldw	5	240
Aras	Rustu Tevfik	Turkish project	no	Sound tape reel. With: Fahrettin Altay.	1	240
Aronson	Albert Henry	Social Security		4 Sound tape reels. With: William Mitchell, Fran	4	61
Ascher	Charles	Biographical Interviews		Sound tape reels.	4	375
Aung	U H'tin	Biographical Interviews		2 sound tape reels.	2	309
Autrey	Herman	Biographical Interviews	no	Sound tape reel.	1	90
Autry	Gene	Hollywood Film Industry Project		1 Sound cassette.	1	49
Baldwin	James	Biographical Interviews		5 Sound tape reels.	5	120
Ball	George W.	Ethnic Groups and American Foreign Policy		3 sound tape reels.	3	78
Barzun	Jacques	Diana Trilling Project		1 Sound tape reel.	1	93
Beauchamp	Jacques A.	Black Journalists Project	no	1 Sound tape reel. With: Robert M. Ratcliffe. A	1	240
Benjamin	Herbert	Biographical Interviews		Sound tape reel.	1	101
Berger	Lawrence	Columbia Crisis of 1968 Project		1 Sound tape reel.	1	100
Berlin	Isaiah	Diana Trilling Project	no	1 Sound tape reel. With: Aline Berlin, John Holla	1	137
Bernard	Viola	Northside Center for Child Development Project	no	2 sound cassettes.	2	144
Bernstein	Bernice	Social Security		1 Sound tape reel. With : J. Douglas Brown.	1	120
Billington	Ray Allen	American Historians		1 Sound tape reel.	1	100

Content Displayed From Columbia's Fedora/Blacklight-based Staff Collection Viewer

Columbia University Libraries Staff Collection Viewer

Log Out [spd1]
Saved Searches Search History
Saved Items

You searched for:

Start over

capra x Projects › Audio Preservation x

Bookmark

Frank Capra Oral History Interview

From: **Rendered Audio** (Click for More Information)

Title:	Frank Capra Oral History Interview
Form:	oral histories
Project:	Preserving Historic Audio Content
CLIO ID:	7024590
Record ID:	http://test.cul.columbia.edu/aggregation/rdf/Capra_Frank_ANDOTHERS_xx-xx-5x_7024590_r
XML Metadata:	MODS, DublinCore

2 Member Items

List

Images Only

More information in CLIO

Title:	Frank Capra Oral History Interview : Session 1, Unknown Date
ID:	http://test.cul.columbia.edu/aggregation /rdf/Capra_Frank_ANDOTHERS_xx-xx-5x_7024590_r_1
Download Metadata:	MODS, DublinCore

MIME Type	Dimensions	File Size	Link
audio/x-wav		540505 Kb	[Download] [DC]
audio/x-wav		1764924 Kb	[Download] [DC]

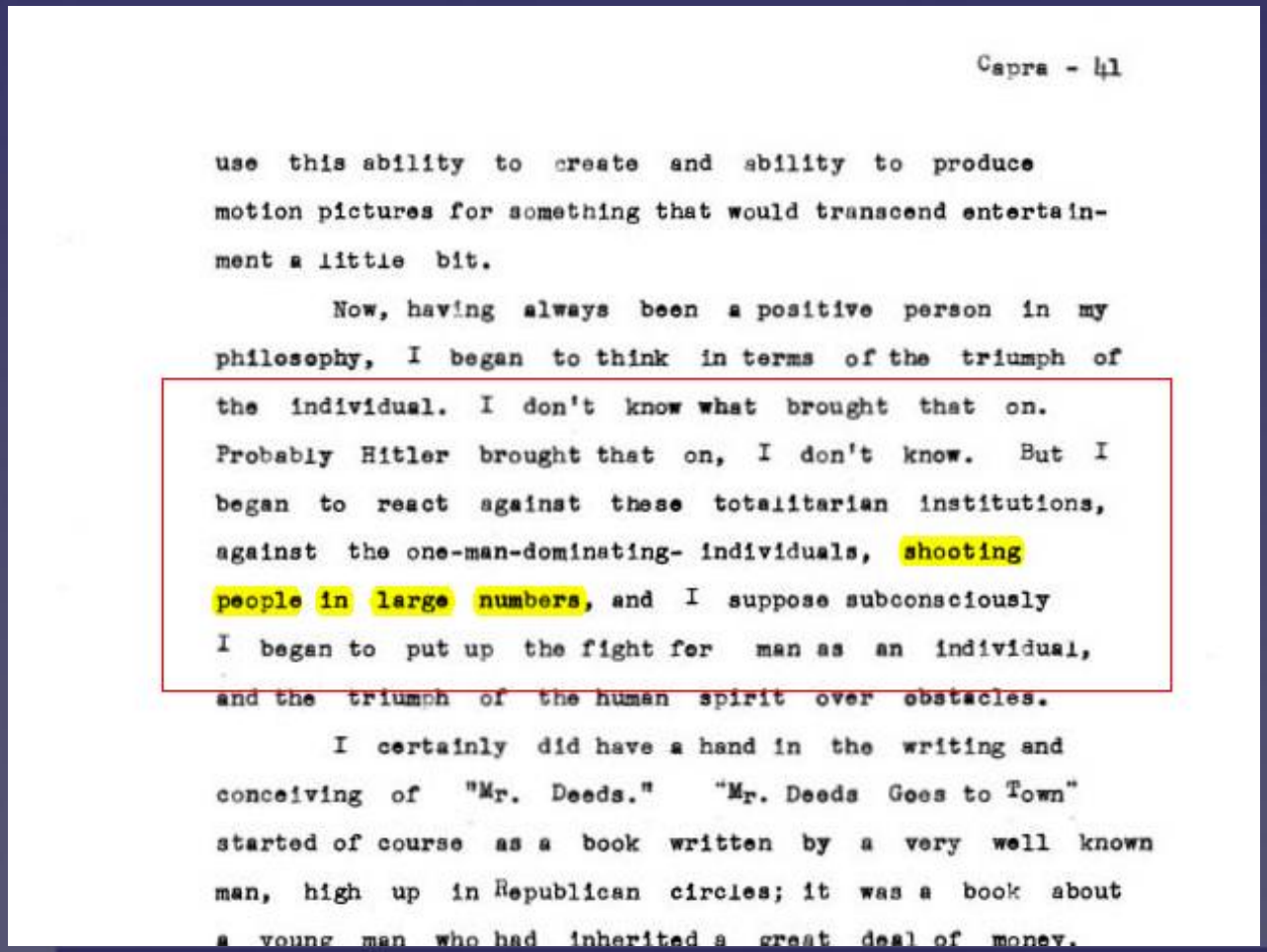
More information in CLIO

Title:	Frank Capra Oral History Interview : Session 2, Unknown Date
ID:	http://test.cul.columbia.edu/aggregation /rdf/Capra_Frank_ANDOTHERS_xx-xx-5x_7024590_r_2
Download Metadata:	MODS, DublinCore

MIME Type	Dimensions	File Size	Link
audio/x-wav		523985 Kb	[Download] [DC]
audio/x-wav		1710982 Kb	[Download] [DC]

Frank Capra (1960), Transcript vs. Audio

(*Noticed when preparing this presentation)*



Capra - 41

use this ability to create and ability to produce motion pictures for something that would transcend entertainment a little bit.

Now, having always been a positive person in my philosophy, I began to think in terms of the triumph of the individual. I don't know what brought that on. Probably Hitler brought that on, I don't know. But I began to react against these totalitarian institutions, against the one-man-dominating- individuals, shooting people in large numbers, and I suppose subconsciously I began to put up the fight for man as an individual, and the triumph of the human spirit over obstacles.

I certainly did have a hand in the writing and conceiving of "Mr. Deeds." "Mr. Deeds Goes to Town" started of course as a book written by a very well known man, high up in Republican circles; it was a book about a young man who had inherited a great deal of money.

Audio version:
“and treating people as numbers” (!)

Issues & Considerations

What did we achieve?

Were there other, simpler ways to do it?

Do we really need all that metadata?

Specific Outcomes #1

- the content is reliably preserved for the future
- the content is preserved as an original "content artifact"
- the content has been reorganized to provide a coherent "interview narrative“
- the process for rendering access files is replicable and correctable

Specific Outcomes #2

- technical ‘provenance’ has been documented
- content can be validated as “authentic”
- content is structured so that it can be managed bibliographically
- content (10 TB) has been ingested into our long-term preservation repository (Fedora)
- content can be made publicly accessible to the extent permitted by author agreements

Strategic Outcomes #1

- Well-developed structural model for digitization of oral history audio
- Fully-developed support by a vendor who can now produce standards-based output
- Solid procedures for local cataloging of analog and digital oral histories

Strategic Outcomes #2

- Built out tools and workflows for ingesting complex content into Fedora repository
- Developed new features for Fedora/Blacklight Staff Collection Viewer to accommodate complex content
- Columbia can share approach with other institutions starting similar projects

Were there other simpler ways to do it?

- Preserve only “content artifacts”
  - Leave providing good access to the future
- Preserve and provide access only to “interview narrative”
  - Do not store, describe or map master files

Right Decisions for This Project?

Our choices for this project did in fact fully meet our goals for preservation and access.

It was also a good choice to build out our existing Fedora / Blacklight and metadata environment rather than develop new, ad hoc approaches for this project.

Did we really need all that metadata?

Only time will tell ….

Questions?