

# Unhiding the audiovisual past at Columbia University Libraries

Violeta Ilik, Head of Digital Collections & Preservation Systems

\*Timothy Ryan Mendenhall, Metadata Librarian

Dina Sokolova, Digital Preservation Librarian

Melanie Wacker, Metadata Coordinator

\*Alex Whelan, Time-Based Media Metadata Librarian

ALA Annual 2019: ALCTS TSWEIG

24 June 2019

# Background

- 2015: Libraries' Strategic Directions (SD) launched, Provost's Advisory Committee on Libraries re-convened
  - Expedited processing/cataloging, improve discovery/access
  - 2016 PACL report: special concerns over audiovisual formats and hidden collections
- 2016-2018: Hidden Collections & Preservation of Time-Based Media initiatives (SD Advance Knowledge)
- 2018-2019: Time-Based Media Pilot Project: Two concurrent phases
  - Previously digitized content unavailable to users (DCG)
  - Columbia Center for Oral History (CCOH)
- 2019-2021: process remaining backlog of analog audiovisual recordings

# Project Scope: Pilot phase, 2018-2019

Comprising special collections from:

- Rare Book and Manuscript Library
- Columbia Center for Oral History
- University Archives
- C.V. Starr East Asian Library
- Avery Fine Art and Architectural Library
- Burke Library at Union Theological Seminary
- Gabe M. Wiener Music & Arts Library
- Lehman Social Sciences Library / Global Studies Division

Material types:

- 10,000<sup>~</sup> physical objects: audio cassettes, audio reels, audio discs, videotapes, films
- 6,500<sup>~</sup> intellectual entities

# Stakeholders

- **Time-Based Media Initiative Steering Committee**
- **Distinctive Collections Group (DCG)**: directors of special collections libraries (CUL)
- **Columbia Center for Oral History (CCOH)**: curatorial, archival processing, metadata remediation and enhancement
- **Preservation and Digital Conversion Division (PDCD)**: digitization preparation, vendor liaison, quality control
- **Digital Library and Scholarly Technology (DLST)**: digital preservation, asset management, software and web development (including homegrown DAMS: Hyacinth)
- **Original and Special Materials Cataloging (OSMC)**: metadata remediation and enhancement
- Additional staff implicated at all special collections units

# Working across departments

## 4.1: Workflow: Previously digitized content (DCG)

TIME-BASED MEDIA DIGITIZATION - ARCHIVAL COLLECTIONS

Violeta Ilik | May 21, 2019

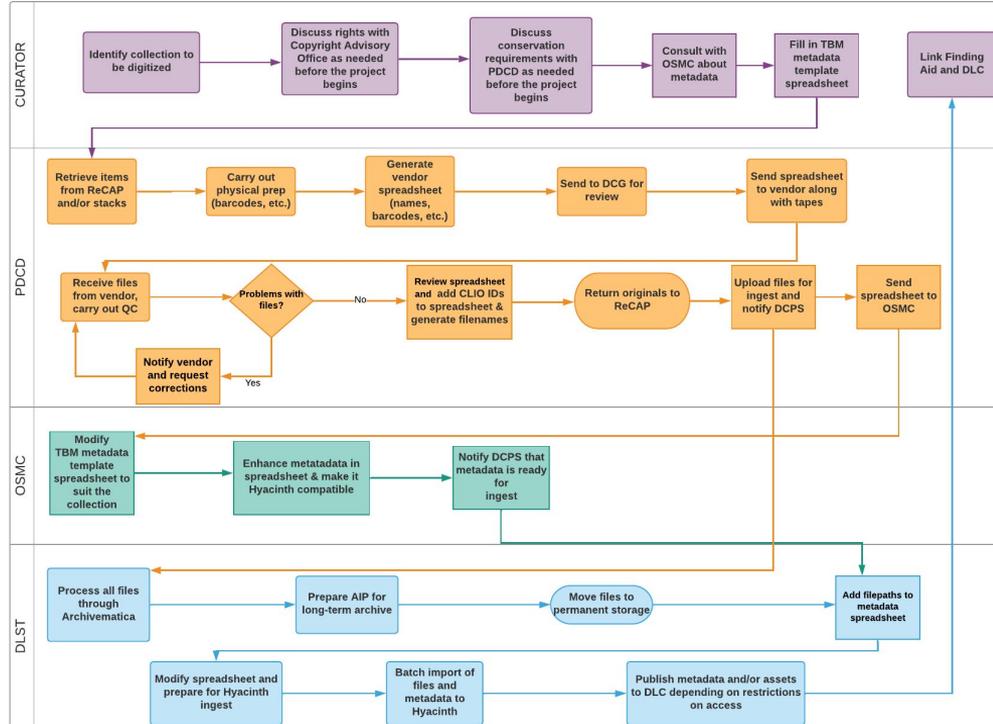
### Departments:

Curator (Rare Book and Manuscript Library etc.)

Preservation & Digital Conversion Division

Original and Special Materials Cataloging

Digital Library & Scholarly Technologies



### Collaboration points:

Evaluating selections and preparing deadlines

Creating or augmenting initial metadata spreadsheet per collection with technical metadata

Remediating and enhancing metadata per specifications for ingest

Unifying descriptive, preservation, and technical metadata; publishing content

# Working across departments

## 4.2 Workflow: Columbia Center for Oral History (CCOH)

### Departments:

Preservation & Digital Conversion Division

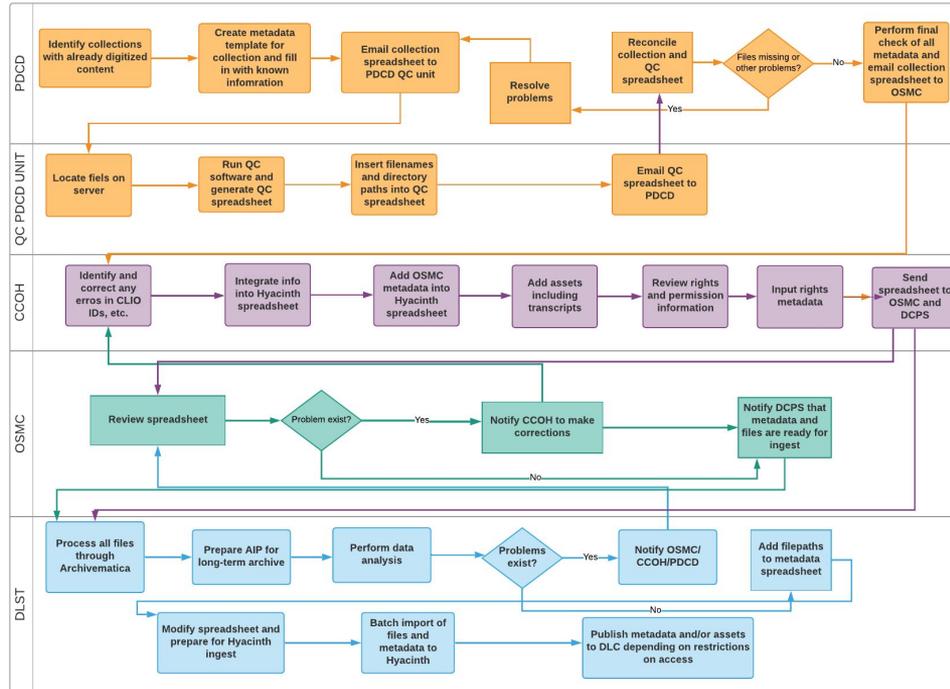
Curator (Oral Histories)

Original and Special Materials Cataloging

Digital Library & Scholarly Technologies

TIME-BASED MEDIA REMEDIATION OF ALREADY DIGITIZED ORAL HISTORY CONTENT

Violeta Ilik | May 21, 2019



### Collaboration points:

Identifying what oral histories had been previously digitized and collating assets (audio, transcripts)

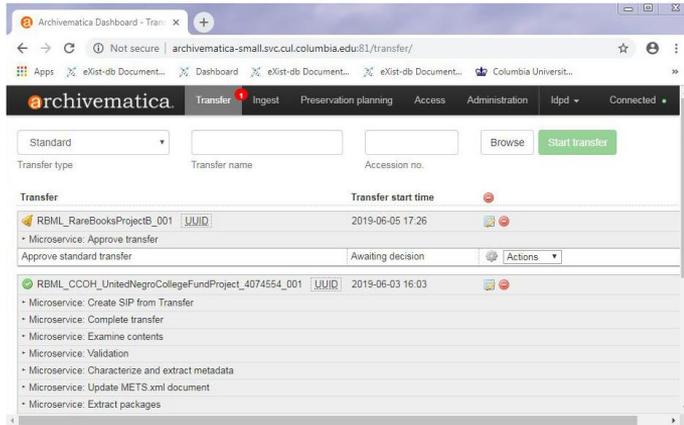
Developing metadata standards for oral histories, remediating legacy records

Refining asset structure to account for legacy practice of recording multiple oral histories onto single master tape

Unifying descriptive, preservation, and technical metadata; syncing ILS and DAMS metadata; publishing content

# Working across departments

## 4.3 Preservation of digital content using Archivematica



- Open-source OAIS-compliant digital preservation system
- Ingests Submission Information Packages (SIPs) and prepares standards-based Archival Information Packages (AIPs) for long-term storage
- Performs automated and customizable digital preservation actions
- Generates METS files containing technical, structural, descriptive, rights, and PREMIS preservation metadata

## 5: Side implications of project

### Accessibility/usability of materials: from separate Carnegie grant project

- Rights module for parsing and displaying restrictions
- Synchronized transcription for oral history audio

### Collection management

- Inventories
- Reference priorities

RIGHTS  STATEMENTS

[Statements](#) [About](#) [Documentation](#) [Get Involved](#) [Blog](#)

#### RIGHTS STATEMENTS

RightsStatements.org currently provides 12 different rights statements that can be used by cultural heritage institutions to communicate the copyright and re-use status of digital objects to the public. The rights statements have been designed with both human users and machine users (such as search engines) in mind and are made available as linked data. Each rights statement is located at a unique URI.

The rights statements have been specifically developed for the needs of cultural heritage institutions and online cultural heritage aggregation platforms and are not intended to be used by individuals to license their own creations. (If you are looking for a tool to license your own works you should consider using one of the [Creative Commons licenses](#)).

#### Three Categories of Rights Statements

IN COPYRIGHT



Statements for works that are in

NO COPYRIGHT



Statements for works that are not in

OTHER



Statements for works where the

## 6: Challenges

### Scale of Operation

- Project planning around earlier steps
- Overlapping project deadlines
- Bottlenecking
- Cross-departmental communication

### Production hurdles

- The nature of audio-visual metadata
  - Working with legacy metadata
  - Creating minimal-level standards
- Changes in requirements
  - File specs (e.g. “rendered” files)
  - Item/asset structure vs. asset-only (regularly enabling new metadata fields)
  - Mapping MARC-to-Hyacinth (homegrown DAM)
- Re-designing digital asset management in real time
  - Hyacinth 3

Field	Hyacinth element name	default mapping	pbooks_and_1 edgers	mapping = oral_history
Primary Name Role (Personal or corporate)	name-1:name_role-1:name_role_term.value	100 or 110 \$e (text) or \$4 (code)	n/a	default
Conference Name Role	name-1:name_role-1:name_role_term.value	711 \$j	n/a	default
Primary Conference Name Role	name-1:name_role-1:name_role_term.value	111 \$j	n/a	default
Primary Name (Corporate)	name-1:name_term.value + name-1:name_term.name_type "corporate" + name-1:name_usage_primary "TRUE"	110 1 or 110 2 subfields \$a, \$b, \$c, \$d, \$g \$n	default	default
Name (Corporate)	name-1:name_term.value + name-1:name_term.name_type "corporate"	710 1, and 710 2. For both, use subfields a,b,c,d,g,n	default	default
Primary Name (Conference/Meeting/Event)	name-1:name_term.value + name-1:name_term.name_type "conference" + name-1:name_usage_primary "TRUE"	111 2 \$a \$c \$d \$e \$ n \$q	n/a	default
Name (Conference/Meeting/Event)	name-1:name_term.value + name-1:name_term.name_type "conference"	711 2 \$a \$c \$d \$e \$ n \$q	n/a	default
Primary Name (Personal)	name-1:name_term.value + name-1:name_term.name_type "personal" + name-1:name_usage_primary "TRUE"	100 1, subfields a,b,d,q	default	default
Name (Personal)	name-1:name_term.value + name-1:name_term.name_type "personal"	700 1 \$a,\$b,\$c,\$d,\$g,\$q	default	default
Note	note-1:note_value	500 \$a	default	default
Biographical note	note-[x]:note_value + note-[x]:note_type="biographical"	545 \$a	n/a	default

# Project Documentation, Communication & Tracking

- Project documentation on the DLST Project Wiki -- including tracking spreadsheet
- Data Dictionary Wiki documenting commonly used elements for audio/visual materials
- Metadata Wiki documenting decisions, files and notes
- Hyacinth Column Headings Wiki containing directions on conversion to ingest format
- OpenRefine Wiki documenting conversion scripts and GREL code snippets
- JIRA for specific technical issues

## Project Documentation, Communication & Tracking

### 4) Tibor Serly Papers (IN HYACINTH)

PRD Metadata for Remediation (posted 08/01/2018)

Remediated metadata (status 08/02/2018)

Issue: The spreadsheet contains multiple file versions (master, service). It is currently unresolved how Hyacinth will handle this. The spreadsheet will need to be adjusted once this has been determined.

Remediated metadata (Updated column headings for file types, 09/07/18)

Remediated metadata completed by DLST (9/17/2018)

AIP file inventory

Content and metadata imported to Hyacinth (09/20/2018). (need permission to see it. Please email Violeta at [vi2143@columbia.edu](mailto:vi2143@columbia.edu) for access)

# Project Statistics & Assessment

TBM Year 1 Collection / Content Status ☆ 🔄

File Edit View Insert Format Data Tools Add-ons Help Last edit was made 3 days ago by Brian Lucero

100% \$ % .0 \_00 123 Arial 10 B I U A 🗑️ 📄 📑 📊 📈 📉 📋 📌 📍 📎 📏 📐 📑 📒 📓 📔 📕 📖 📗 📘 📙 📚 📛 📜 📝 📞 📟 📠 📡 📢 📣 📤 📥 📦 📧 📨 📩 📪 📫 📬 📭 📮 📯 📰 📱 📲 📳 📴 📵 📶 📷 📸 📹 📺 📻 📼 📽 📾 📿 📠 📡 📢 📣 📤 📥 📦 📧 📨 📩 📪 📫 📬 📭 📮 📯 📰 📱 📲 📳 📴 📵 📶 📷 📸 📹 📺 📻 📼 📿

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Name of Collection / Content (with link to CLIO collection record)	Custodial Division	PDCD Spreadsheet Posted (with link)	Number of physical objects	Number of Items	Number of Master Files	Reviewed by Dina (with date if available)	OSMC Revised Spreadsheet Posted (with link)	Content Processed through Archivematica (with link to inventory)	Content ingested into Hyacinth (with Hyacinth link)	Curatorial review of metadata / rights / presentation	Content published to DLC (staging)	Content integrated into finding aid
2	<b>Totals</b>			4166	3178	6044							
3	<a href="#">Abe Kobo Collection</a>	Starr	<a href="#">11/12/2018</a>	1	1	2	11/13/2018	<a href="#">12/06/2018</a>	<a href="#">11/16/2018</a>	<a href="#">12/14/2018</a>		<a href="#">12/14/2018</a>	
4	<a href="#">Abraham Anderson interviews with Diana Trilling</a>	RBML		4	1	6							
5	<a href="#">Albert Ellis Papers</a>	RBML	<a href="#">01/19/2019</a>	1	1	2	01/29/2018	<a href="#">01/29/2018</a>	<a href="#">1/29/2019</a>	<a href="#">01/31/2019</a>		<a href="#">01/31/2019</a>	
6	<a href="#">Albert Goldman Papers</a>	RBML	<a href="#">11/12/2018</a>	207	181	324	11/13/2018	<a href="#">02/19/2019</a>	<a href="#">12/6/2018</a>	<a href="#">02/27/2019</a>		<a href="#">02/27/2019</a>	
7	<a href="#">Albert Goldman Papers Pt. 2</a>	RBML	<a href="#">01/03/2019</a>	5	5	6	01/04/2019	<a href="#">04/16/2019</a>	<a href="#">1/10/2016</a>	<a href="#">04/19/2019</a>		<a href="#">4/19/2019</a>	
8	<a href="#">American Bureau for Medical Aid to China Records</a>	RBML	<a href="#">02/19/2019</a>	28	22	38	03/27/2019	<a href="#">05/07/2019</a>					
9	<a href="#">Amin Baraka Papers</a>	RBML	<a href="#">09/26/2018</a>	225	194	379	09/26/2018	<a href="#">10/24/2018</a>	<a href="#">10/10/2018</a>	<a href="#">11/14/2018</a>		<a href="#">12/14/2018</a>	
10	<a href="#">Amos Vogel Papers</a>	RBML	<a href="#">11/12/2018</a>	1	1	1	11/13/2018	<a href="#">12/4/2018</a>	<a href="#">11/16/2018</a>	<a href="#">12/14/2018</a>		<a href="#">12/14/2018</a>	
11	<a href="#">Andrew W. Cordier Papers</a>	RBML	<a href="#">10/01/2018</a>	51	38	63	10/01/2018	<a href="#">11/27/2018</a>	<a href="#">10/09/2018</a>	<a href="#">12/03/2018</a>		<a href="#">12/14/2018</a>	
12	<a href="#">Angelika Balabanova Sound Tapes</a>	RBML	<a href="#">07/24/2018</a>	11	5	11	yes	<a href="#">08/02/2018</a>	<a href="#">09/12/2018</a>	<a href="#">10/25/2018</a>		<a href="#">12/14/2018</a>	
13	<a href="#">Arthur Mitchell Collection</a>	RBML	<a href="#">11/12/2018</a>	43	41	43	11/13/2018	<a href="#">04/16/2019</a>	<a href="#">02/01/2019</a>	<a href="#">4/19/2019</a>		<a href="#">4/19/2019</a>	
14	<a href="#">Bella Abzug Papers</a>	RBML	<a href="#">01/08/2019</a>	42	40	55	4/16/2019	<a href="#">05/01/2019</a>					
15	<a href="#">Barbara Curtis Adachi Bunraku Collection</a>	Starr	<a href="#">11/12/2018</a>	7	7	14	11/13/2018	<a href="#">12/10/2018</a>	<a href="#">11/16/2018</a>	<a href="#">01/10/2019</a>		<a href="#">01/10/2019</a>	
16	<a href="#">Barney Rosset Papers</a>	RBML	<a href="#">11/12/2018</a>	1	1	1	11/13/2018	<a href="#">12/12/2018</a>	<a href="#">11/16/2018</a>	<a href="#">01/17/2019</a>		<a href="#">01/17/2019</a>	
17	<b>Bob Fass (TBM)</b>	RBML	part of the new digitization project; no remediation spreadsheet will be submitted										
18	<a href="#">C. L. R. James Papers</a>	RBML	<a href="#">11/12/2018</a>	6	6	10	11/13/2018	<a href="#">12/17/2018</a>	<a href="#">11/16/2018</a>	<a href="#">01/17/2019</a>		<a href="#">01/17/2019</a>	
19	<a href="#">Chaliapin Family Papers</a>	RBML	<a href="#">11/12/2018</a>	1	1	1	11/13/2018	<a href="#">12/26/2018</a>	<a href="#">11/16/2018</a>	<a href="#">1/2/2019</a>		<a href="#">01/02/2019</a>	
20	<a href="#">Chinese Independent Film Archive: Zhen Cui</a>	Starr	<a href="#">09/06/2018</a>	2	1	2	09/06/2018	<a href="#">09/07/2018</a>	<a href="#">09/27/2018</a>	<a href="#">10/23/2018</a>		<a href="#">12/14/2018</a>	
21	<a href="#">Cecile Starr Papers</a>	RBML	<a href="#">11/12/2018</a>	1	1	2	11/13/2018	<a href="#">12/12/2018</a>	<a href="#">11/16/2018</a>	<a href="#">01/17/2019</a>		<a href="#">01/17/2019</a>	
22	<a href="#">Columbia Armenian Archive</a>	RBML	<a href="#">12/14/2018</a>	1	1	1	12/20/2018	Will be processed through OH workflow	<a href="#">01/10/2019</a>				
23	<a href="#">Columbia University historical recordings collection</a>	RBML	<a href="#">01/08/2019</a>	8	8	12	01/16/2019	<a href="#">01/31/2019</a>	<a href="#">2/15/2019</a>	<a href="#">02/27/2019</a>		<a href="#">02/27/2019</a>	
24	<a href="#">Columbia University Media Collection</a>	Univ Arch	<a href="#">12/14/2018</a>	2	2	3	01/02/2019	<a href="#">01/03/2019</a>	<a href="#">1/15/2019</a>	<a href="#">01/22/2019</a>		<a href="#">01/22/2019</a>	
25	<a href="#">Columbian Library</a>	Univ Arch	<a href="#">12/17/2018</a>	1	1	2	01/02/2019	<a href="#">03/25/2019</a>	<a href="#">01/15/2019</a>	<a href="#">04/17/2019</a>		<a href="#">04/17/2019</a>	
26	<a href="#">Conant Audio-Visual Collection</a>	Starr	<a href="#">02/14/2019</a>	83	74	87							
27	<a href="#">Composers Forum Concerts</a>	Music	<a href="#">02/04/2019</a>	451	159	458			<a href="#">3/29/2019</a>				
28	<a href="#">David Dinkins papers</a>	RBML	<a href="#">10/01/2018</a>	2	2	2	10/01/2018	<a href="#">10/04/2018</a>	<a href="#">10/26/2018</a>	<a href="#">11/01/2018</a>		<a href="#">12/14/2018</a>	
29	<a href="#">David Hamburg Papers</a>	RBML	<a href="#">10/11/2018</a>	20	7	32	10/11/2018	<a href="#">10/18/2018</a>	<a href="#">10/29/2018</a>	<a href="#">11/14/2018</a>		<a href="#">12/14/2018</a>	
30	<a href="#">Department of Intercollegiate Athletics Records</a>	Univ Arch	<a href="#">11/30/2018</a>	1	1	1	12/03/2018	<a href="#">12/06/2018</a>	<a href="#">12/03/2018</a>	<a href="#">12/14/2018</a>		<a href="#">12/14/2018</a>	

# Unhiding the past

## 8: Examples of finished collections/items/assets

Showing results for: Library Collection > Arthur Mitchell Collection

- 10 Mexico-Holland-etc  
Mitchell, Arthur, 1934-2018
- 11 London I  
Mitchell, Arthur, 1934-2018
- 16 New York City Ballet  
Mitchell, Arthur, 1934-2018
- 38 Spoleto-Me Dancing  
Mitchell, Arthur, 1934-2018
- 45 Shook  
Mitchell, Arthur, 1934-2018
- 6 Luchino-Spoleto  
Mitchell, Arthur, 1934-2018
- 9 Manila-Indiana  
Mitchell, Arthur, 1934-2018
- Agon  
Mitchell, Arthur, 1934-2018

Showing results for: Formal Genre > transcripts

- 21 transcript
- 22 transcript
- 23 transcript
- 24 transcript
- 25 transcript
- 26 transcript
- 27 transcript
- 28 transcript

Showing results for: Language > French

- 1 Call Machine Takeover: Amen...
- 2 Oubli Horizon
- 3 Condi
- 4 Condi - Hémémathoron, Lire...
- 5 Condi - Siegau, Lire de la Se...
- 6 Condi du no 2
- 7 Daniel Maxmin interview with ...
- 8 Daniel Maxmin, L'Isle Sobel...

# Outcomes, lessons learned, future directions

## Phase 2 (Oct. 2019 - ): Even More Materials

- Workflows and documentation are stable, but perhaps not final
- Adding positions, but also adding production specs
  - Current charter cites 60,000 additional items to be processed 2019-2021

## Future Directions

- Archives Space Integration
  - Clean-up of legacy finding aids
  - Linking between Archives Space and digital assets/metadata
- Project management division

# Questions?



Violeta Ilik: [vi2143@columbia.edu](mailto:vi2143@columbia.edu),

\*Timothy Ryan Mendenhall: [trm2151@columbia.edu](mailto:trm2151@columbia.edu)

Dina Sokolova: [ds2057@columbia.edu](mailto:ds2057@columbia.edu)

Melanie Wacker: [mw2064@columbia.edu](mailto:mw2064@columbia.edu)

\*Alex Whelan: [aw3195@columbia.edu](mailto:aw3195@columbia.edu)