Search and You Shall Find: Making Oral Histories Searchable with Transcripts and Metadata

Amanda Hurford
Digital Initiatives Multimedia Developer
AAHurford@bsu.edu

Maren Read
Assistant Archivist for Manuscript Collections
MLRead@bsu.edu
Why is searching valuable?

- Find what you’re looking for
  - Empower your audience to do more than just browse
- Opening resources to wider audience
  - OAI
  - Google
- Organize the collection
How to make oral histories searchable?

- Transcripts
- Metadata

Archives and Special Collections
University Libraries

Title
Interviewee
Date Recorded
Run Time
Location Recorded
Subject
Media Type
Original Physical Format
Project Identification
Archival Identification
Digital Identification
Repository
Collection
Subcollection
Ordering Information
Copyright

Abram, Dr. Sam 02-12-2003 (Interview)
Kraemer, Anne
Moore, Ashley
Anderson, Michelle
Abram, Sam
2003-02-12
1:25:51
Muncie, Indiana
Oral history
Muncie (Ind.) -- Ethnic relations
African Americans -- Indiana -- Muncie -- History
African Americans -- Indiana -- Muncie -- Biography
African Americans -- Indiana -- Muncie -- Interviews
Streaming Audio (WMA)
Audio Cassette
Getting a Living - M03GL10-I-02/12
R251-5
OSM_003.wma
Ball State University Libraries, Archives and Special Collections
Middletown Digital Oral History Project Collection
Other Side of Middletown Collection
To order a copy, inquire about permissions, or for information about the oral history, please call 765-285-7108.

[Text of interview]

Archivist: When do you remember to start?

Interviewee: Whenever (laughter)

Archivist: So this subject of our interview is, now?

Interviewee: Training the young, educating as well as informed indoctrination of children. Now children learn how to act and have not to act. It's important because when we want to focus on our chapter.

Archivist: Well let me read next and you'll see all around us now and you will use education because I mean--there's a program we give you to follow on either. Let me pass you a book. I think it meets with this house and, ah, we are taking their messages of educating the young, cheer, it's not until there. One of the challenges in the house weren't other additional help as those early years.
Transcripts

Planning

1. Choose to transcribe
2. Choose your transcribers
3. Choose your tools
Choose to Transcribe

- Advantages of Full-Text Transcripts
  - Access
    - Searchability
    - Hearing/Visually impaired
  - Ease
    - Assessing relevance
    - Understanding speakers
  - Preservation
    - Backup of audio information
...or Not to Transcribe

- Other options
  - Abstracts
  - Indexes
  - Audio only

- What to do with existing transcripts?
  - Scan?
  - OCR?
  - Retype?
Goodall: Is that who you’re trying to think of?

Wright: It was.

Goodall: Oh, uh-huh.

Wright: Oh, I’m awful forgetful.

White: Do you remember how you got the job working for them?

Wright: Well, I was working on Jackson Street for an old lady, school teacher, and this woman come past there—rich people, but I can’t call her—she come past there and asked who I was and she says, says, “Oh, Ida Wright. She’s a wonderful worker.” I can’t recall her name. And she says, “Why don’t you work for me?” So she come down by our house, we were living over on 8th Street. She come down there one day and asked if I would work for her and I worked for her for fourteen years. Worked for twenty-two years. And the last place I worked was for Mrs., her husband just died.

Goodall: How did you get back and forth out there? Just ride the streetcar or walk—

Wright: the streetcar over there went way on. See, that’s when I froze my feet standing along waiting on the streetcar.

Goodall: How many hours did you usually have to work at that time?

Wright: I’d go to work at five, about 4:30.

Goodall: In the morning?
Choose Your Transcribers

- **In-House**
  - **Pros:**
    - Quality control
    - Train transcribers
    - Access to verification resources
  - **Cons:**
    - Time
    - Staff
    - Training

- **Commercial Service**
  - **Pros:**
    - Time
    - Experience
  - **Cons:**
    - Little/no quality control
    - Cost
    - Lack of subject/term knowledge
    - Paperwork/Oversight needed
Choose your tools

- OCR Software
- Scanners
- Voice Recognition Software
- Transcription Software
  - Express Scribe
  - Start Stop
  - Windows Media Player
  - USB Foot Pedal
- What we chose
  - Transcription Software: Express Scribe
Getting Started

- Decisions
  - Format and Style
    - Baylor Guide
  - What to leave in, what to take out
    - Privacy
    - Speech patterns and variations
    - Editorial comments
  - Time Codes
  - Names List
  - Equipment (Software, Headphones, Foot pedals)
Training

- Who should transcribe? Who should edit?
- Training should provide:
  - Instructional materials (style guide, wiki)
  - Dos and don’ts
  - FAQs
  - Background information on interviews
  - Demonstration of equipment
- Tips
  - If possible, gather in one place at one time
  - Make it hands on – let them practice
  - Keep it simple
Work Flow

- Splitting tasks
  - Transcribe
  - Edit
  - Quality Control
  - Package and Finalize
- Evaluating Progress
- Documenting Progress
  - Shared spreadsheets
  - Wiki
Oral Histories

Contents

1. Scope
2. Transcription
   2.1 Supplemental Transcription Information
   2.2 Common Capitalization and Spelling Questions
   2.3 Layout Guide
3. Digitization
4. Status
5. Metadata

Scope

This project is LSTA grant funded and aimed at digitizing and transcribing the Middletown Digital Oral History Collection. It is presented online through the Digital Media Repository here: [http://lib.bsu.edu/MdOrHist/midorhist.php](http://lib.bsu.edu/MdOrHist/midorhist.php). It consists of the following subcollections:

- Black Muncie History Project - 23 interviews
- Black Middletown - 35 interviews
- Other Side of Middletown - 56 interviews
- Middletown Jewish Oral History Project I - 21 interviews
- Middletown Jewish Oral History Project II - 24 interviews
- Muncie Catholic Churches Oral History Project - 30 interviews
  1. St. Francis of Assisi Catholic Church Oral History Project
  2. St. Lawrence Catholic Church Oral History Project
  3. St. Mary Catholic Church Oral History Project

Transcription

We are following the Baylor University Institute for Oral History Style Guide for guidance in transcribing our oral histories.

Supplemental Transcription Information

For most situations the Baylor Style Guide will suffice as our authoritative guide; however, we have decided to make a few minor changes and additions to facilitate our project and to clarify some unaddressed finer points.

Please alert the rest of the project’s staff to any major changes.

Abbreviating Titles: If written in conjunction with any part of the person’s name (i.e. first name, initials, or surname) use an abbreviated title.
Presentation

- Transcript format
- Arrangement in CONTENTdm
  - Separate items rather than compound object
  - Custom thumbnails
- Linking to get audio and transcript together
- Presentation and searchability
<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Interviewer</th>
<th>Date Recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Abram, Dr. Sam 02-12-2003 (Interview Transcript)</td>
<td>Kraemer, Anne ; Moore, Ashley ; Anderson, Michelle</td>
<td>2003-02-12</td>
</tr>
<tr>
<td>2</td>
<td>Abram, Dr. Sam 02-12-2003 (Interview)</td>
<td>Kraemer, Anne ; Moore, Ashley ; Anderson, Michelle</td>
<td>2003-02-12</td>
</tr>
<tr>
<td>3</td>
<td>Abram, Dr. Sam 02-21-2003 (Interview Transcript)</td>
<td>Burke, Carla ; Kissel, Carrie</td>
<td>2003-02-21</td>
</tr>
<tr>
<td>4</td>
<td>Abram, Dr. Sam 02-21-2003 (Interview)</td>
<td>Burke, Carla ; Kissel, Carrie</td>
<td>2003-02-21</td>
</tr>
<tr>
<td>5</td>
<td>Abram, Pastor Renzie 01-17-2003 (Interview Transcript)</td>
<td>Booth, Jessica; Efaw, Eric; Gawlowski, Daniel;</td>
<td>2003-01-17</td>
</tr>
</tbody>
</table>
Metadata

- Structured description (unlike transcripts)
- Functions of metadata
  - Find
  - Identify
  - Select
  - Obtain
- Dublin Core
- Sources of Information
Metadata: Planning

- Who will create metadata?
- When should you create metadata?
- What level of metadata?
- What Dublin Core elements will you use?
- How will you label those fields?

- Our fields…
<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>Title</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Link to Transcript</td>
<td>Relation-Is Version Of</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Link to Audio</td>
<td>Relation-Is Version Of</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Interviewer</td>
<td>Creator</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Interviewee</td>
<td>Creator</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>Date Recorded</td>
<td>Date-Created</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>See Photo of Interviewee</td>
<td>Relation</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Run Time</td>
<td>Format-Extent</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>Location Recorded</td>
<td>Coverage-Spatial</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Subject</td>
<td>Subject</td>
<td>Text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>Notes</td>
<td>Description</td>
<td>Text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>Media Type</td>
<td>Type</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>Original Physical Format</td>
<td>Format</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>Project Identification</td>
<td>Identifier</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>Archival Identification</td>
<td>Identifier</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>Digital Identification</td>
<td>Identifier</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>17</td>
<td>Repository</td>
<td>Source</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>Collection</td>
<td>Relation-Is Part Of</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>Subcollection</td>
<td>Relation-Is Part Of</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>Ordering Information</td>
<td>Rights</td>
<td>Text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>21</td>
<td>Copyright</td>
<td>Rights</td>
<td>Text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>Full Text</td>
<td>Description</td>
<td>Full Text Search</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Lessons Learned

- Prepare collections
- TIME!!!
- PEOPLE!!!
- Name list
- Subject headings