

# The SpokenWeb Project: Documenting Humanities-Oriented Spoken Collections

Jason Camlot (Concordia U), Michael O'Driscoll (U Alberta), Sean Luyk (U Alberta), Holly Pickering (U Alberta), Ali Azarpanah (U Alberta), Geoffrey Rockwell (U Alberta)

1. Jason Camlot: What is SpokenWeb?

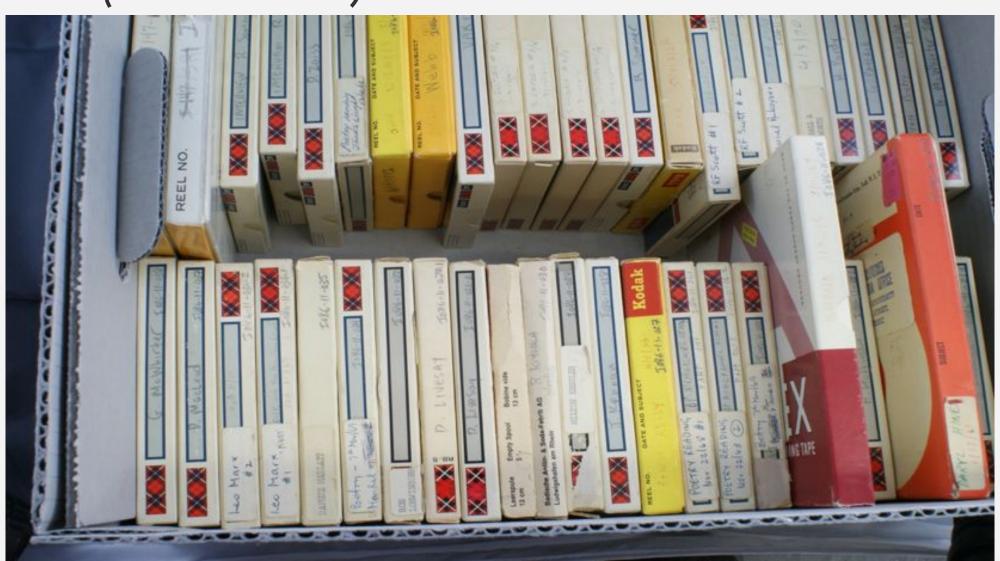
What is digital about SpokenWeb?

Media Formats, Audiotextual Forms, Literary Affordances

Digital projects of the SpokenWeb research program

What is SpokenWeb?

Reel to Reel tapes documenting the Sir George Williams Poetry Series (1965-1974)



### What is The SpokenWeb?

- Aims, methods, activities: In what ways will it engage with sound?
- SpokenWeb Collections: The (expanding) scope of our audio corpus.

### What is SpokenWeb?

A SSHRC-funded research partnership (PG) and network currently made up of 15+ partners (institutional, community, organizational) across the US and Canada and 50+ researchers (and 30+ students and PDFS) working on the development of digitized collections of literary sound recordings in order to make them accessible for use in research and teaching, and on research, teaching and "knowledge mobilization" activities with the digitized collections.

### Who is SpokenWeb?

### **Present Institutional Partners**

Holding Collections: Simon Fraser U, U British Columbia Okanagan, U Alberta, U Calgary, U Toronto, Victoria U in the U Toronto, Concordia U.

Holding Relevant Expertise: U Texas at Austin, UC Davis, NYU, U Pennsylvania (PennSound), U Victoria, etc.

### **Other Partners**

Wired on Words, Archive MTL, Blue Metropolis Foundation, Compute Canada Researchers, selected (Co-Applicants and Collaborators)

<u>Lit</u>: Jason Camlot, Gregory Betts, Clint Burnham, Al Filreis, Michelle Levy, Smaro Kamboureli, Mike O'Driscoll, Karis Shearer, Darren Wershler, Jason Wiens, Cecily Devereux, Carole Gerson, Nick Mount.

<u>Library and Archives</u>: Rebecca Dowson, Roma Kail, Sean Luyk, Loryl MacDonald, Tomasz Neugebauer, Felicity Tayler, Jared Wiercinski, Geoff Harder, Marjorie Mitchell, Tony Power, David Seubert.

Community: Ian Ferrier, Louis Rastelli, William St-Hillaire.

<u>DH</u>: Tanya Clement, Adam Hammond, Marit MacArthur, Brian McFee, Lee Miller (as consultant), Geoffrey Rockwell, Miles Thorogood.

Oral History: Steven High.

Media History: Patrick Feaster, Matthew Rubery, Jenterey Sayers, Jonathan Sterne.

Podcasting: Hannah McGregor.

Student Researchers, PDFs and "HQP" (lots of them). A Student Advisory Committee of approx. 25 students to start with.

### Some Aims

- 1) Inventory and index existing collections in Canada, held by institutional and community partners, to establish a broad understanding of our national holdings of literary sound recordings, and, digitally preserve and aggregate these collections by developing a metadata scheme.
- 2) Study literary recordings using interdisciplinary theories and methods of "audiotextual criticism".
- 3) Develop and test computational methods for analyzing literary sound.
- 4) Create online modules, interactive aural exhibitions and new works of sound art from the digitized archive for use in research, teaching and public presentation, allowing the archive to speak and tell stories in new ways.

### Methods

The SW research program is designed to explore

- 1) the history of literary performance and practice as discernible in sound recording media,
- 2) cultural effects of the circulation of digitized literary recordings, and
- 3) how engagement with sound recordings can have the effect of changing disciplinary practices in research and teaching, by mixing methodologies from literary, media and sound studies, oral history, digital humanities, digital libraries, and archival theory.

We'll introduce new methodological approaches to the field of literary studies, revise the nature and format of literary canons, etc.

# SpokenWeb Collections: The (expanding) scope of our audio corpus

Time frame: mainly 1960-2000

**SFU** collections hundreds of hours of readings, lectures, interviews, literary conversations, **UBCO** collection holds multiple homemade literary audio tape collections of Canadian poets including readings, compilations and literary dictation.

U Alberta collections document local reading series.

**U Calgary** collections hold recordings made by individual Canadian authors.

**Concordia U** collections hold documentary recordings of literary performance, multimedia literary/art practice, individual author tape collections

**U Vic** holds a rich collection of formal literary interviews

**U Toronto** holds collections that document practices of literary pedagogy, and French language reading series.

**Wired on Words** holds a "words and music" reading series that has run from 1980s to the present.

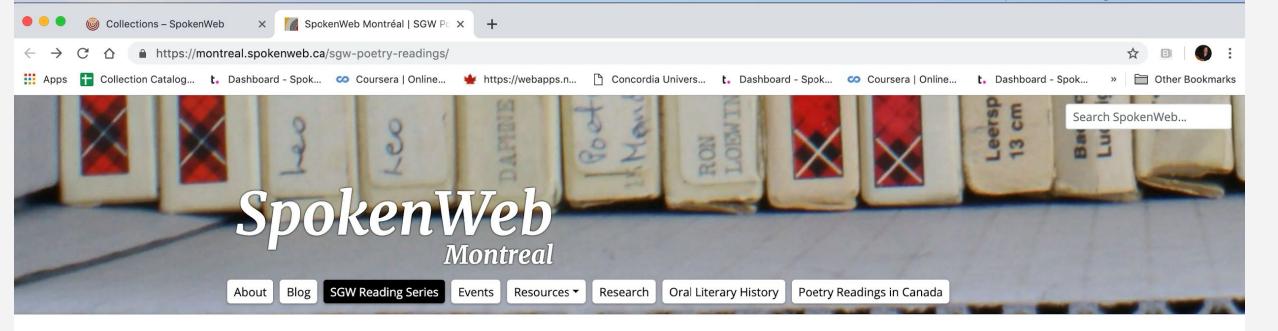
ARCMTL counterculture collections, dumpster dive materials.

**Hidden Collections:** We have a mandate to seek out collections that have not been deposited in archives, digitize, preserve and integrate them into our project (with the source materials to be deposited in partner archives or kept by the holder of the collection.

A few examples:

**Yellow Door Readings** (36 Cassette tapes documenting a readings series from the early 1990s). **Ultimatum Festival and Readings**: 39 reel to reel tapes plus several U-Matic video tapes of a bilingual punk poetry series that took place in Montreal in the mid 1980s.

We did a survey and there are MANY more such collections out there in shoeboxes, in basements.



#### SGW POETRY READING SERIES

Between 1965 and 1974 members of the Sir George Williams University (SGWU, now Concordia University) English Department in Montreal hosted a series of poetry readings that was conceived as an ongoing encounter between local poets and the avant-garde poetics of some of the most important writers from the United States and the rest of Canada. Sponsored by "The Poetry Committee" of the Faculty of Arts and the SGWU English Department—and organized primarily by English professors Howard Fink, Stanton Hoffman, Wynn Francis, Irving Layton, Roy Kiyooka, and (from 1967-71) George Bowering—these readings involved more than sixty poets from across North America.

Known simply as "The Poetry Series", audio recordings of these readings were made on Mylar 1 mil. tape using mobile reel-to-reel tape machines. The Concordia University Archives received a grant in 2007 that has allowed all 65 reels of tape (more than 100 hours of audio) to be digitized. So, this sound from an interesting period of transformation in Canadian poetics, and of self-scrutiny for Montreal poetry, represents a rich and useable archive for scholarly research.



#### **FILTERS**

Search









What is digital about SpokenWeb?

Researchers pursue projects via collaborative Task Forces.

### (META) DATA PROCESSING AND MANAGEMENT

- Metadata Task Force
- Systems Task Force
- RDM Task Force
- Analog/Digitization Task Force
- Rights Management Task Force

#### **ANALYSIS**

- Audio Signal Analysis Task Force
   CONTEXTUALIZATION/HISTORICIZATION
- Oral History Task Force

### KNOWLEDGE MOBILIZATION

- Podcasting Task Force
- Pedagogy Task Force
- Archival Events Task Force
- Community Collections Task Force

Literary or Audiotextual Research doesn't have a task force as it is an

assumed form of activity, being reshaped by the rest...

### ETC.

**LOCATION** 

PRESERVATION/ DIGITIZATION

**DESCRIPTION** 

DATA-IZATION

**ANALYSIS** 

PRODUCTION/ CREATION

ACCESS/
DISSEMINATION

**CIRCULATION** 





# Spoken West: The UAlberta Collection

Michael O'Driscoll, UAlberta



# A brief and heroic history!



| Single T        |         |  | Track Bobine No.           |
|-----------------|---------|--|----------------------------|
| Date<br>Recorde | d Time  | SUBJECT  | Track No. Speed Footage    |
|                 |         | DEPT OF ENGLISH POETRY REPOLDS   | 1 33,456                   |
|                 |         | James Reaney.  | Jan 16 1969                |
|                 |         | in a reddition of a Suribune ode   |                            |
|                 | undog"! |  | The Wald of Childhad .     |
| 249 0           | 16" Li  | the Orphan annu 1 225 "Trip up o<br>ammoth Cheece / 273 Roch group with  | home made instrument       |
| 371 "5          | The Hum | ming 3 ind " / 389 "The Killdeer" /<br>Poems (with visuals on board)   |                            |
| Date            | Durée   | SUJET  | Piste No. Vitesse Lonqueur |
|                 | REPLACI | EFECTIVE IN LABELLING, PACKAGING OR MANUFACT<br>ED. THERE ARE NO OTHER WARRANTIES EXPRESSED OF<br>AMER MAGNETICS | OR IMPLIED.                |

# Reading the audio object



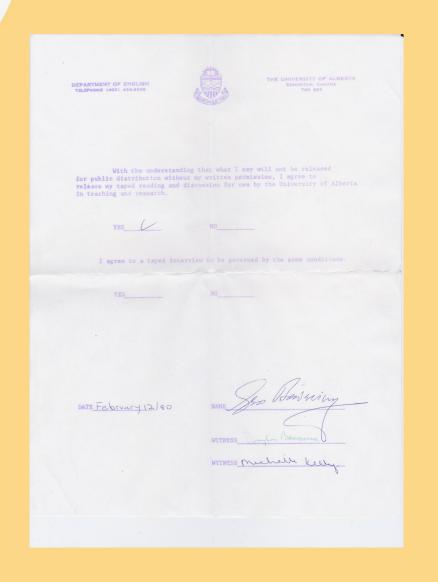
Close listening as a practice of consignation

# Extra-poetic speech as information field

... or, Phyllis Webb: "I have to catch my breath."

# Oral history: archive stories

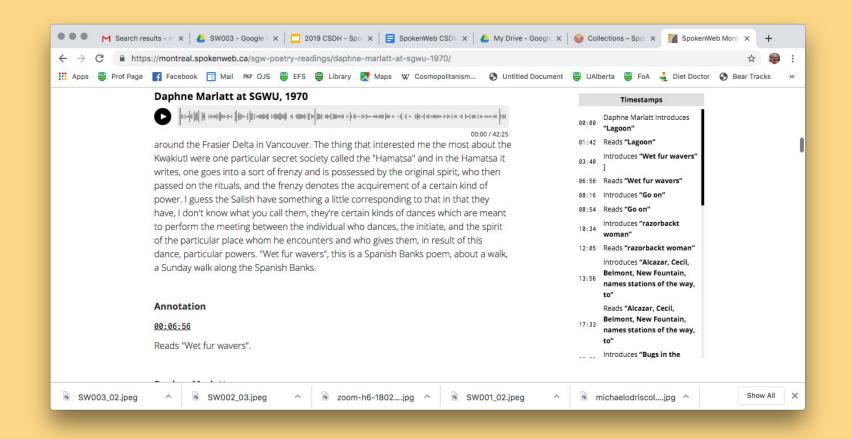




# Archival history and the paratextual environment

Audiographic codes and the production of meaning





# Indexing and durational media

# Conclusions

mo@ualberta.ca



# Metadata/Systems Considerations

Sean Luyk, UAlberta





https://era-av.library.ualberta.ca



https://www.avalonmediasystem.org

ERA Home

education & research archive

luyk@ualberta.ca 🕶

enter search terms

Search ERA A+V

▶ AUDIO + VIDEO

Manage Content Manage Groups Manage Selected Items (0) Home Browse Playlists Contact

| Browse By        |       |
|------------------|-------|
| Format           | ~     |
| Sound Recording  | 1,763 |
| Moving Image     | 516   |
| Main contributor | >     |
| Date             | >     |
| Genres           | >     |
| Collection       | >     |
| Language         | >     |

| Limit to       |   |  |  |  |
|----------------|---|--|--|--|
| Published      | > |  |  |  |
| Created by     | > |  |  |  |
| External Group | > |  |  |  |
| Date Digitized | > |  |  |  |
| Date Ingested  | > |  |  |  |

« Previous | 1 - 10 of 2,287 | Next »

Sort by Date -

10 per page -

☐ Select All

☐ Select

☐ Select

☐ Select

□ Select

1. Ivars Taurins Lecture & Rehearsal with USO, Madrigal Singers & Concert Choir (2:00:10)

Date: {2003-03-26, 2003-03-27}



Main contributors: Taurins, Ivars

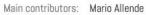
2. Student Composer Recital - 260 & 460 levels (1:07:03)

Date: {1995-04-03, 1995-04-05}

Main contributors: Various

3. Programa Nosotros. (59:43)

Date: {1993-02-21, 1993-02-28} (Creation date: 1993-02-24)



Summary: Continuation of a series. A conference in Vancouver with the consulate General of Peru Jaime Pomerada on promoting Latin-American culture, and the difficulties of living far from

one's home and cul...

4. Programa Nosotros. Bafochi (1:02:13)

Date: {1991-10-20, 1991-10-27} (Creation date: 1991-10-20)

Main contributors: Sergio Contreras

Summary: This program presents a report on the relief telethon for the cholera

epidemic in Peru. The report discussed the donations gathered and the various people and groups involved in making the telethon...

**>** audio + video

# ERA A+V/Avalon: Core Features

- Collection search and discovery
- Media deposit workflow management
- Single sign on (through campus computing ID)
- Learning Tool Integration with UA LMS
- Basic access control
- Ability to embed content into websites, create playlists, and display closed captioning
- Standards-based metadata (MODS), persistent URLs, and structural metadata (basic XML using div, span, begin, end elements)

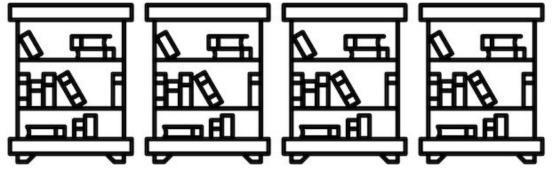
### Outline

- 1. High-level metadata considerations
- 2. SpokenWeb Metadata Task Force
- 3. High-level systems considerations

## Collections as Data



### **Always Already Computational**



https://collectionsasdata.github.io/

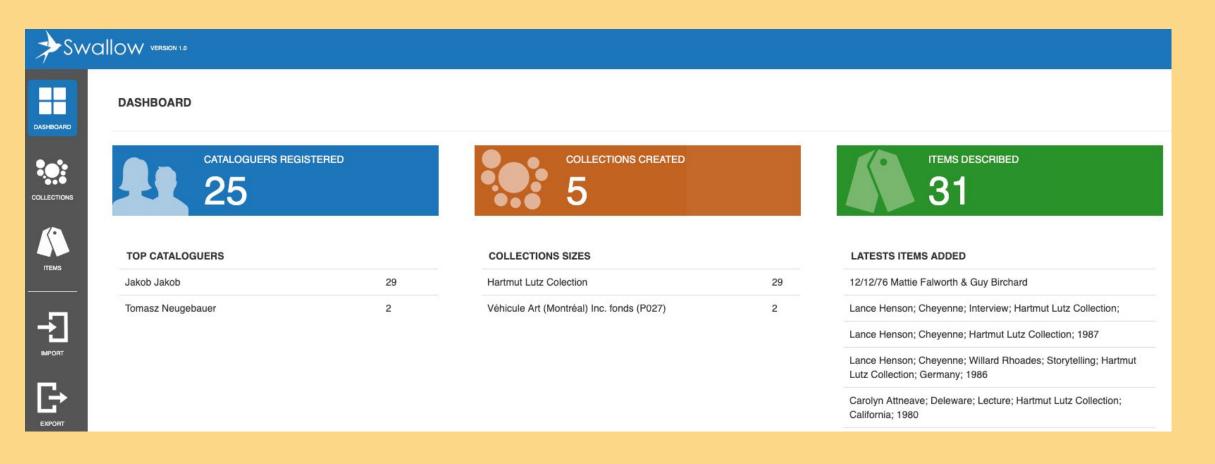
"Audiovisual materials are notoriously under represented in digital humanities precisely because they often lack the detailed data (or metadata) that supports their effective discovery, identification, and use by researchers, students, instructors, or collections staff"

-Tanya Clement, "Unsolved Problems in the Humanities Data Generation Workflow: Digitization Complexities, Undiscoverable Audiovisual Materials, and Limited Training for Information Professionals," Always Already Computational: Collections as Data Final Report, p. 106.

"...even the most basic metadata is often more useful for understanding a collection than many librarians imagine. Simply having author or creator information, or language information, can be very helpful"

-Miriam Posner, "Actually Useful Collection Data: Some Infrastructure Suggestions," Always Already Computational: Collections as Data Final Report, p. 135.

## Metadata Task Force

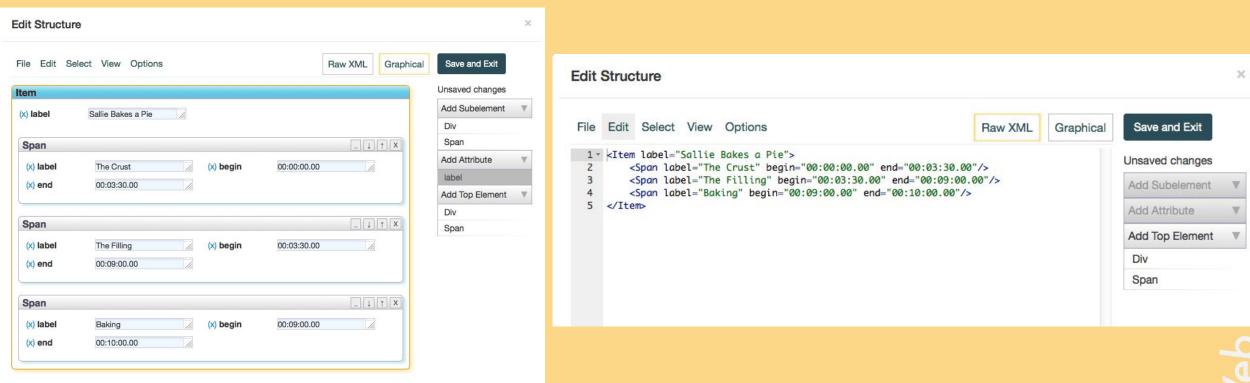


Tomasz Neugebauer, Francisco Berrizbeitia, and Jared Wiercinski. Swallow (SpokenWeb Audio Metadata Ingest System) v. 1.0, Landing Page.



# SpokenWeb

## Structural Metadata in Avalon



Avalon Media System, Collection Manager's Guide, "Adding Structure to Files Using the Graphical XML Editor" <a href="https://wiki.dlib.indiana.edu/display/VarVideo/Adding+Structure+to+Files+Using+the+Graphical+XML+Editor">https://wiki.dlib.indiana.edu/display/VarVideo/Adding+Structure+to+Files+Using+the+Graphical+XML+Editor</a>



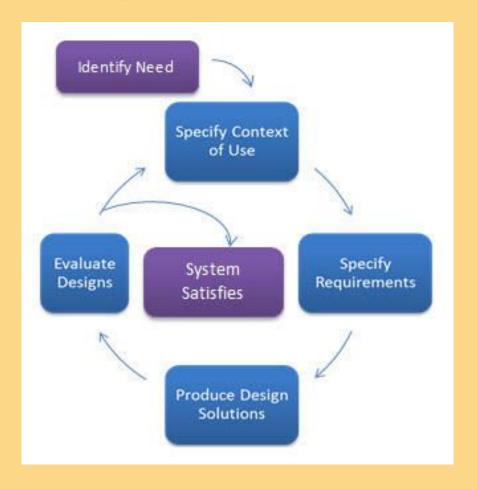
## Structural Metadata - New Graphical Editor



# Sterne's Roadmap vs. ERA A+V/Avalon

| Functions   | Current Functionality  |
|---|------------------------|
| Platform independence   | ×                      |
| Limited downgrading   | <b>✓</b>               |
| Ease of navigation  | <b>✓</b>               |
| Use of lossless formats   | 🗶 (at ingest only)     |
| Supports looping  | ×                      |
| Displays related files with audio (e.g. images)                           | ×                      |
| Range of visualization options (waveform, spectral display, rhythm, etc.) | ×                      |
| Ability to configure audio (e.g. mono vs. stereo)                         | ×                      |
| Manipulate audio (speed up, slow down, scrubbing, freezing, marking)      | ✓ (scrubbing, marking) |

## Systems Task Force - Next Steps



- Gather Requirements:
  - Specify the context of use
  - Determine user requirements
  - Recommend design solutions
  - Evaluate designs through usability testing
- Iterate changes as necessary

"Scholarly software designers should more consider the research on user-centered design approaches...and theories associated with user experience"

-Gibbs, Fred and Trevor Owens. "Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs." *Digital Humanities Quarterly* 6/2 (2012) <a href="http://www.digitalhumanities.org/dhg/vol/6/2/000136/000136.html">http://www.digitalhumanities.org/dhg/vol/6/2/000136/000136.html</a>



#### sean.luyk@ualberta.ca



# Metadata for the Spoken Word

Holly Pickering, UAlberta



#### Avalon

#### Metadata fields:

- Title (Required)
- Date Issued (Required)
- Bibliographic ID
- Bibliographic ID Label
- Other Identifier
- Other Identifier Type
- Creator
- Contributor
- Genre(s)
- Publisher(s)
- Date Created
- Abstract
- Language(s)

- Physical Description
- Related Item URL
- Related Item Label
- Topical Subject
- Geographic Subject
- Temporal Subject
- Terms of Use
- Table of Contents
- Statement of Responsibility
- Note
- Note Type



#### Metadata fields:

- Cataloguer Last Name, First Name
- Partner Institution
- Contributing Unit
- Source Collection
- Source Collection Description
- Source Collection ID
- Item ID Number / Identifier
- Persistent URL (if available)
- Asset Image File Name
- Recording Type
- AV Type
- Material Designation
- Physical Composition
- Storage Capacity of Artifact
- Extent Side
- Playing Speed
- Track Configuration
- Playback Mode
- Equalization
- Tape Brand
- Accompanying Material
- Other physical description

- Generations
- Sound Quality
- Physical condition
- Conservation
- Other Digitized file name
- Digital File Path
- Channels
- Sample rate
- Precision
- Duration THH:MM:SS
- ❖ Size
- ❖ Bit rate
- Encoding
- Title
- Title Note
- Title Source
- Language
- Production Context
- Genre
- Creator 1
- Creator URL

- Creator Role 1
- Contributor 1 Last, First Name
- Contributor 1 URL
- Contributor Role 1
- Contributor 2
- Contributor 2 URL
- Contributor Role 2
- Creator/Contributor Notes
- Date
- Date Type
- Date Source(s) of Information
- Date Notes
- Address
- Venue (enter if available)
- Address Notes
- Latitude
- Longitude
- URI
- Rights
- Rights Note
- Time Stamp Begin

- Time Stamp End
- Duration
- Title of Work
- Other Notes





### SpokenWeb UAlberta

#### Metadata fields:

- SW Catalogue Number
- AV Catalogue Number (original)
- Format
- Reader(s)
- Additional Speaker(s)
- Genre
- Venue
- Date

- Technical Specifications
- Recording Specifications
- Photo/Scan Files
- Notes from reel/cassette case and/or accompanying documents
- Time Stamps
- General note



#### SpokenWeb UAlberta and Avalon

https://docs.google.com/spreadsheets/d/1m3upspTHNH8Om2oF4xYp JRxNoDITXr7vkuqTjKRxCHk/edit?usp=sharing



#### SpokenWeb UAlberta and SpokenWeb

https://docs.google.com/spreadsheets/d/1FzZDmHEUqpDsI3CcN0DUC qqx0157FDEjfqMpTiBuXFw/edit?usp=sharing



### SpokenWeb UAlberta and Avalon and SpokenWeb

https://docs.google.com/spreadsheets/d/1jgN6zs8xu 04rK9E1CkQg 18PYLfqRJ4qxJaaGRZ8NSI/edit?usp=sharing



#### Resultant Recommendations

- Cataloguer name
  - This comes from SpokenWeb's schema. The cataloguer is an important part of the collection of these items.
- Source Collection
  - > This comes from SpokenWeb's schema. Our collection should have a name.
- Source Collection Description
  - This comes from SpokenWeb's schema. Our collection should have a description.
- Other Identifier Type
  - This comes from Avalon we have a Catalogue Number ("Other Identifier" in Avalon-speak) but we need to assign it a type for Avalon.
- Digital File Path
  - This comes from SpokenWeb's schema. Helps with organizing.
- ❖ Title
  - This comes from both SpokenWeb and Avalon Avalon requires this.
- ❖ Title Source
  - This comes from SpokenWeb's schema. Where did this title come from? Is it from the box or from the cataloguer?
- Language
  - From both SpokenWeb and Avalon. Easy enough to do and helpful for searches.
- Topical Subject
  - This comes from Avalon These come from the Library of Congress Subject Headings and are important for people searching a site to be able to find what they're looking for. Keywords, basically.



#### Resultant Recommendations 2

- Statement of Responsibility
  - From both SpokenWeb and Avalon. This tells us who is who (creator, reader, publisher, etc. as creator or contributor) for Avalon's system.
- Date Type
  - From SpokenWeb. What kind of date is our date? We will be using multiple types of dates in Avalon likely, so it would be good to have it written down what they are.
- Terms of Use
  - From both SpokenWeb and Avalon. Crucial for putting things online.
- Time Stamp Document, found here:

https://docs.google.com/spreadsheets/d/1HRrCTjUNK1LX75ocHkqlbC8yGQqhUM0GRauH-tS5sRQ/edit?usp=sharing

- ➤ Includes:
  - Unique ID
  - Time In
  - Time Out
  - Title
  - Text
  - Notes
  - Type
- Related Item URL
  - From Avalon. Nice to have for an online database
- Related Item Label
  - > From Avalon. What is this item?

#### Future Steps

- mandatory versus optional categories
- Ul design according to metadata categories



## SpokenWeb.ca





## Interfacing with the Spoken Word

Ali Azarpanah, UAlberta



### Why a University of Alberta portal?

- The SpokenWeb project is a partnership of distributed centers
- Intellectual property issues
- The need for a family resemblance to the central site

#### How do we design a portal to the spoken word?

The characteristics of the persona/scenario/wireframe process

- Controlled method
- Time-consuming method
- Series of distinct phases

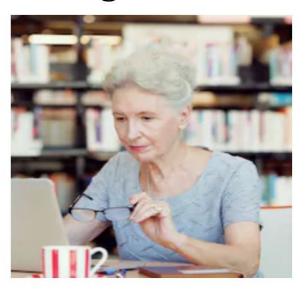
#### Personas

"Personas are not real people, but they represent them throughout the design process. They are hypothetical archetypes of actual users. Although they are imaginary, they are defined with significant rigor and precision" (Alan Cooper)

"Not knowing who you are designing for increases the risk of ending up with a product that makes nobody sufficiently satisfied" (Stephanidis, Universal Access in Human-Computer Interaction)

#### Personas

#### 1. Abigail McGlenn



Occupation: retired teacher

Age: 73

**Location: Fredericton** 

Potential audience group: Researcher

Abigail McGlenn is a retired teacher. She interested in poems. Every week she goes to a book club where she and her friends read a book in the field of Literature. She is not an expert but she wants to become more informed. She is enthusiastically searching the portal and listening to the tapes. She enjoys listening the tapes and reading the texts; even, she posts SpokenWeb portal's materials on her social media.

#### Scenario

### Scenarios "explain the future activities", they depicts steps that a user performs to complete the task.

- 1. Abigail and her friends read a short story from Jorge Luis Borges in their book club.
- 2. She searches his name to know more about him.
- 3. She finds Borges's page on the SpokenWeb portal.
- 4. She listens to the tape and reads the text.
- 5. She takes some notes.
- 6. She bookmarks the page.
- 7. She posts the link on her book club's Facebook page.

#### Scenario

#### Website user flow

Abigail McGlenn (Persona 1, Retired Teacher)

Homepage



Search bar



Collection



Tape

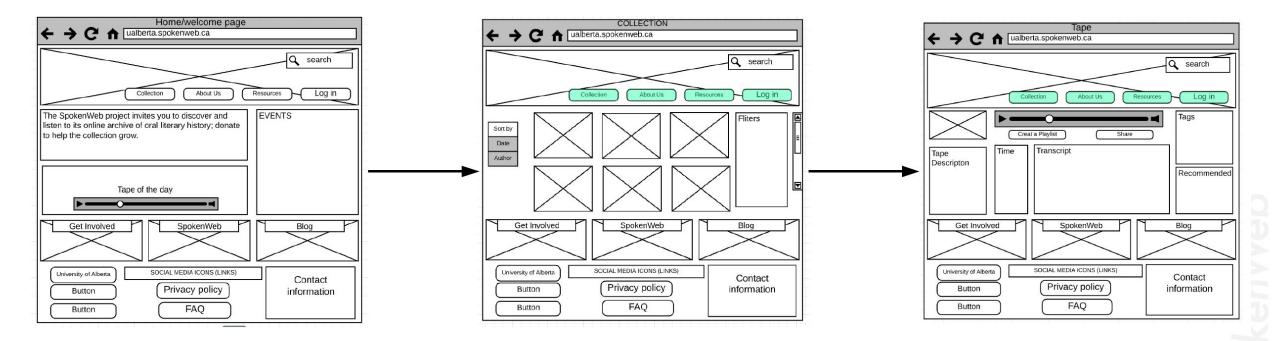


Sharing on social media



#### Wireframes

"Wireframes are schematic presentations that define a webpage's content and functionality structure in order to portray the page concept before it is designed and developed" (Kotze & et al.)





#### Implementation framework

- ERA AV (Education and Research Archive Audio-Visual) for the Content
- Catfish for the Interface

A **Web Content Management System (WCMS)** "provides website authoring, collaboration, and administration tools designed to allow users with little knowledge of web programming languages or markup languages to create and manage website content with relative ease" (Jin & Lin, Advances in Electronic Commerce).

https://arc.arts.ualberta.ca/



Thank you

Discussion / Questions

spokenweb.ca



## SpokenWeb.ca

