Developing Open Oral Histories in Islandora

Lingling Jiang @sprklinginfo
Kim Pham @tolloid
Kirsta Stapelfeldt @kstapelfeldt
Edward Garrett @digitalutsc

Digital Scholarship Unit
Overview

● Who are we?
● Unique joint partnership
● Oral Histories Solution Pack
● Live demo
Digital Scholarship Unit

Build capacity for digital scholarship through shared infrastructure and OS components
Nearby Studies is a joint program developed between the Departments of Historical and Cultural Studies, Women's and Gender Studies, and Geography. It aims to focus on local issues affecting Scarborough and the eastern GTA by working with local community groups, historical associations, and other nearby partners.

Dr. Christine Berkowitz & Partners
“The Sherpa project is part of the Eastern Himalaya Research Network, which is an international, multifaceted initiative around research, teaching and learning about the Eastern Himalayas. The Sherpa project partners with communities in the Himalayas for the preservation of Sherpa culture and heritage through oral histories.”
Shockingly Similar Requirements

- Preserve video/audio long-term
- Play video/audio with time-coded info
- Let people collaborate
Pinedrop - Transcripts in Drupal + Solr
Partner - Pinedrop

“Pinedrop creates websites and collaborative research tools for scholarly communities and archives.”

Pinedrop is creating a suite of Drupal modules to help researchers at the University of Virginia to store, search, and display a collection of transcripts which have been time-aligned to media files.

http://pinedrop.com/
Unique joint partnership

Edward Garrett
@Pinedrop

Dr. Chris berkowitz
@UTSC

Dr. Jo Sharma
@UTSC

UTSC Library
DSU
Use existing SPs as dependencies

This module requires the following modules/libraries:

- Islandora
- Tuque
- Islandora Solr Search
- Islandora Video Solution Pack
- Islandora Audio Solution Pack
- Transcripts UI

https://github.com/digitalutsc/islandora_solution_pack_oralhistories
New Oral Histories Content Model

- Contains relevant datastreams from audio/video SP (depending on the source file)
- Contains new TRANSCRIPT & MEDIATRACK datastreams
<table>
<thead>
<tr>
<th>TN</th>
<th>RELS-EXT</th>
<th>Fedora Reserved Datastreams</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>MODS</td>
<td>Form-defined Descriptive Metadata</td>
</tr>
<tr>
<td>OBJ (audio or video)</td>
<td>MP4 (video only)</td>
<td>Derivatives determined by ingest mimetype</td>
</tr>
<tr>
<td>PROXY_MP3 (audio only)</td>
<td>TECHMD</td>
<td>New Datastreams supporting time-encoded information</td>
</tr>
<tr>
<td>TRANSCRIPT</td>
<td>MEDIATRACK</td>
<td></td>
</tr>
</tbody>
</table>
New Datastream - TRANSCRIPT

A custom, extensible, simple XML file for transcript and annotation tiers (and/or more tiers)

Benefits:

- Flexibility and extensibility
- Easy for solr index: A simple xslt can be developed to have GSearch to index each cue.
<?xml version="1.0" encoding="UTF-8"?>
<cues>
<!-- If the entire transcript has one speaker only, use 'solespeaker' element. -->
<!-- Then skip 'speaker' element in 'cue' element level. But DO NOT use them in both places. -->
<!-- If 'solespeaker' element presents in the document, following 'speaker' elements will be ignored. -->
<solespeaker>One Speaker</solespeaker>
<speaker>Different Speaker</speaker>
<!-- 'start' and 'end' elements are start time and end time in seconds for the cue. -->
<start>0.000</start>
<end>12.124</end>
<!-- 'transcript' and/or 'annotation' are default content tiers of the cue. -->
<!-- Extra tier(s) can be added as long as they are listed in the configuration page. -->
<!-- 'transcript' element is required if 'Enable captions/subtitles display' is configured to be true. -->
<transcript>This is the transcript text content.</transcript>
<annotation>This is the annotation content.</annotation>
</cue>

<!-- add more cues with above structure. -->
</cues>
New Datastream - MEDIATRACK

WEBVTT
An emerging simple text based format for displaying timed text tracks for HTML5 <track> element.
It provides captions/subtitles/metadata that is time-aligned with audio or video content.

http://dev.w3.org/html5/webvtt/
WEBVTT

00:00.040 --> 00:02.503
Anna: A rocket is a launch vehicle.

00:02.503 --> 00:07.474
Anna: It can also be a space capsule, usually in the form of a cylinder or

00:07.474 --> 00:10.577
Anna: a tubelike structure that launches from the Earth.</transcript>

00:10.577 --> 00:13.480
Anna: And it can be used to carry humans into space.

00:13.480 --> 00:17.017
Anna: It can also be used to carry supplies such as cargo,

00:17.017 --> 00:25.926
Anna: food, materials or anything that can be carried to the International Space Station and to work up there.

00:25.926 --> 00:30.264
Anna: It can also be used as something to launch satellites into the orbit

00:30.264 --> 00:34.835
Anna: that the military may use or that NASA may use, or it can even be a satellite.
New Configuration Options

- user-defined “layers” or “tiers” element from TRANSCRIPT datastream
- option to create WebVTT from required “transcript” tier
Future Steps

- Support in-browser, user-authored & edited annotations/transcriptions!
- Convert and support transcripts produced by other popular tools?
- Abstract annotations into separate objects to comply with Open Annotation Collaboration Model?
Contact us

- Lingling Jiang, Applications Developer, University of Toronto Scarborough, @sprklinginfo
- Kim Pham, Digital Projects & Technologies Librarian, University of Toronto Scarborough, @tolloid
- Kirsta Stapelfeldt, Coordinator, Digital Scholarship Unit, University of Toronto Scarborough, @kstapelfeldt
- Edward Garrett, Drupal Developer, Pinedrop, edward@pinedrop.com

- Digital Scholarship Unit: digitalscholarship@utsc.utoronto.ca