VIDEO & METADATA (@ CORNELL)

Jason Kovari
jak473@cornell.edu
PBCore 2.0: Public Broadcasting Metadata Dictionary Project

- Developed by the Corporation for Public Broadcasting
- Highly lauded
- Includes descriptive & technical elements
- Complex / well-designed [from my peripheral view]
PBCore 2.0: Public Broadcasting Metadata Dictionary Project

- Well-suited for segment description
  - `<pbcorePart>`
    - Must contain: `pbcoreIdentifier`; `pbcoreTitle`; `pbcoreDescription`
    - May contain: `startTime`; `endTime`; `timeAnnotation`
    - Optional:
      - `pbcoreAssetDate`
      - `pbcoreSubject`
      - `pbcoreGenre`
      - `pbcoreRelation`
      - `pbcoreCreator`
      - `pbcoreRightsSummary`
      - Etc...
VIDEO TECHNICAL METADATA

- Provisioned in PBCore
- VideoMD as technical metadata standard

- Important for rendering, migration and preservation
Cornell’s Video Projects: Collections – Some Examples

Müller-Kluge
- Interviews with Heiner Müller, German playwright
- Well-chaptered & described
- German and English
- http://muller-kluge.library.cornell.edu/ 

Experimental Television Center
- Video Art
- Multiple, (ir-&related) pieces per recording
- Legacy metadata available but problematic
KALTURA
OPEN SOURCE VIDEO STREAMING PLATFORM

- Campus-wide deployment
  - CUL
  - CornellCast (cornell.edu/video/)
  - Blackboard (CIT - Academic Technologies)
  - Video on Demand (CIT)
  - Lab of Ornithology

- Others to come...
KALTURA: ADVANTAGES

- User-generated upload and metadata capture
- Aggregated content and metadata
CORNELL VIDEO METADATA: ELEMENTS OF CONCERN

- Distributed
  - Campus, library & non-Cornell platforms

- Structural Elements
  - Time codes / chaptering

- In-depth segment-level description

- Caption/Transcription (in multiple languages)
  - Current player can switch languages
KALTURA
OPEN SOURCE VIDEO STREAMING PLATFORM

- Library provided metadata field-sets:
  - Cornell-wide baseline
  - CUL extended

- Field-set written into .xsd
FIELD-SET: BASELINE

<Title / >
<Description / >
<Categories / >*
<Keywords / >
<CreationDate / >
<Creator / >
  <CreatorLastName / >
  <CreatorFirstName / >
  <CreatorRole / >*
  <CreatorNetID />
<ContributingUnit / >*
<Language / >*
<Rights / >

*Controlled Field
**FIELD-SET: CUL extended**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;Identifier /&gt;</code></td>
<td><code>&lt;ContainerText /&gt;</code></td>
</tr>
<tr>
<td><code>&lt;AlternativeTitle /&gt;</code></td>
<td><code>&lt;DigitizationSteps /&gt;</code></td>
</tr>
<tr>
<td><code>&lt;Subject /&gt;</code></td>
<td><code>&lt;HardwareSoftware /&gt;</code></td>
</tr>
<tr>
<td><code>&lt;Location /&gt;</code></td>
<td><code>&lt;MasterFileFormat /&gt;</code></td>
</tr>
<tr>
<td><code>&lt;OriginalFormat /&gt;</code></td>
<td><code>&lt;DerivativeCopy /&gt;</code></td>
</tr>
<tr>
<td><code>&lt;SoundType /&gt;</code></td>
<td></td>
</tr>
<tr>
<td><code>&lt;ColorType /&gt;</code></td>
<td><code>&lt;ChapterDescription&gt;</code></td>
</tr>
<tr>
<td><code>&lt;RecordingNote /&gt;</code></td>
<td><code>&lt;CaptionURL&gt;</code></td>
</tr>
<tr>
<td><code>&lt;Collection /&gt;</code></td>
<td><code>&lt;TimeCodeURL&gt;</code></td>
</tr>
<tr>
<td><code>&lt;AccessionNumber /&gt;</code></td>
<td></td>
</tr>
<tr>
<td><code>&lt;Repository /&gt;</code></td>
<td></td>
</tr>
</tbody>
</table>
FIELD-SET: AUTO-CAPTURED

<Bitrate />  <Thumbnail />
<Framerate />  <AssetInternalFlavorsURLs />
<Filesize />  <VideoDownloadURL />
<VideoDuration />  <AudioDownloadURL />
<AudioDuration />
<Width />
<Height />
<FileFormat />
<SourceCodec />
<UploadDate />
<LastModifiedDateTime />
CONTROLLED VOCABULARIES

- Kaltura CV – [not easily extensible]
  - Only controlled in baseline fields

- `<CreatorRole />`
- `<Categories />`
- `<Language />`

- `<ContributingUnit />`
  - Kaltura – VIVO – Kuali
**Kaltura:**

**Metadata Concerns**

- **Nested Elements**
  - unable in Kaltura, currently
    - `<CreatorLastName>` `<CreatorFirstName>` `<CreatorRole>` `<CreatorNetID>` example

- **Non-Repeatable**

- **Still unsure of segment-level metadata**

- **No underlying *common* schema**
  - .xsd per instance
What is Needed (Metadata-Wise)?

- Multi-level / structural description
- Provisioning for multi-language metadata
- Hierarchical metadata
- Repeatable fields
- Fuller-able segment-level metadata
GOOD NEWS

- Simpler projects work well, metadata-wise
  - Instruction videos, non-segmented, non-captioned etc.

- Metadata will allow for categorization on LibeCast-esque site

- Issues of segmentation metadata and captions are being solved
THANK YOU!

Metadata Working Group
2012-May-08

Jason Kovari
jak473@cornell.edu