EVIA Digital Archive New Tools

William G. Cowan

Mike Durbin

Digital Library Program

EVIA Digital Archive

DLP Brown Bag

20 September 2006

EVIA Digital Archive
New Tools

EVIA Digital Archive Development Team:
William Cowan
Nick Hansen
Mike Durbin

EVIA Digital Archive Background

- Mellon Foundation funded Project
  - Planning Phase 2001 – 2002
  - Development Phase 2003 – 2005
  - Sustainability Phase 2006 - 2009
- Pilot Project to digitize video from Ethnomusicologists
  - Extensive use of video for field collecting
  - Desire to preserve this work
  - Provide ability to annotate this work

EVIA Digital Archive Partners

- Partnership between Indiana University and University of Michigan

- University of Michigan – Responsible for digitizing video

- Indiana University – Responsible for Application Development

EVIA Digital Archive Units Involved at IU

- Department of Folklore and Ethnomusicology
  - Ruth Stone
- Archives of Traditional Music
  - Daniel Reed, Alan Burdette, Suzanne Mudge, Mike Casey
- Digital Library Program
  - Jon Dunn, Will Cowan, Nick Hansen, Mike Durbin, Jenn Riley
- UITS Digital Media Network Services
  - James McGookey

EVIA Digital Archive Process

- Transfer video to Digital Betacam tape
- Encode video to MPEG-2 file at 50 Megabits/second
- Archive MPEG-2 file to IU Massive Data Storage System
- Generate derivatives (MPEG-4, H.264) from MPEG-2 file for streaming access
- Annotate video
- Provide access to video and annotations through Web interface

Creating the Digital Archive – Some Statistics

- Size of 50 Mb Digital Master File
  - 10 - 15 gigs per hour of video
  - 100 – 200 gigs per collection
  - 2 – 3 Terabytes per Summer Institute
- Effort to Create File for Application Use
  - 8+ hours per hour of video
  - Includes: dubbing digibeta tape, MPEG2 encoding, checksum generation, transmission and transcoding

EVIA Digital Archive Summer Institute 2006

- From June 19 to June 30, 2006
- Over a dozen Ethnomusicologists from around the world
- Over 150 hours of new digital content
- New Content from:
  - China
  - Kuwait
  - Malawi
  - Tanzania
  - Northern Ireland
  - Mexico
  - Macedonia
  - even Indiana

EVIA Digital Archives Tools

- Search and Browse Web Interface

- Controlled Vocabulary Maintenance

- Video Segmentation and Annotation

Application Development Environment

- Java using Eclipse
- Relational Database for Controlled Vocabulary
- Lucene for indexing
- FEDORA Repository
- Search and Browse – current releases of various browsers
- Uses QuickTime and QuickTime for Java

Search & Browse Interface

- Keyword Search

- Advanced Search

- Results Display / Manage Results

- Video Playback

  - Planned changes to annotation display

  - Planned changes for dock-able components

Search & Browse Interface

- [Demo of Search & Browse]

EVIA Digital Archive Controlled Vocabulary

- Prior to the Summer Institute
  - Ethnomusicologists and cataloguers define categories
    - For EVIA: instruments, languages, social and cultural groups, venues, geographic locations
  - Authority sources chosen by cataloguers
    - Examples: Library of Congress Subject Headings, Getty Thesaurus of Geographic Names

EVIA Digital Archive Controlled Vocabulary

- For each Summer Institute
  - Terms added to categories based on collections' content
  - As needed, new values added by catalogers during video segmentation process
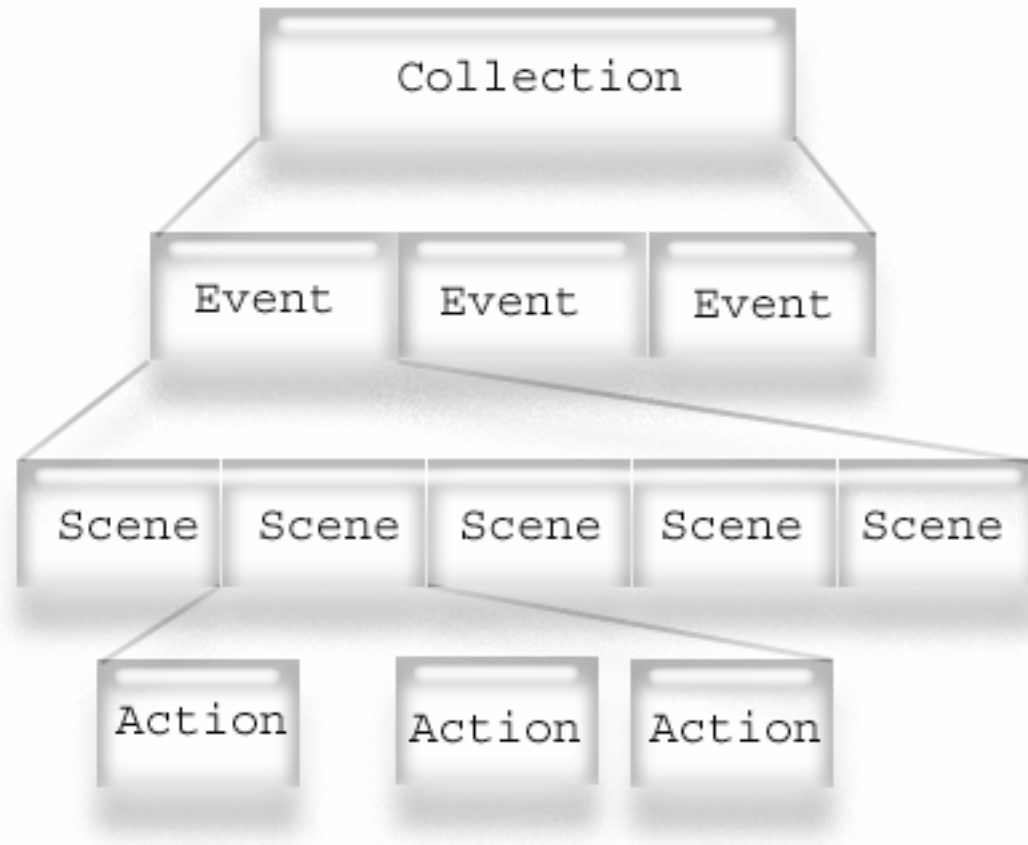
EVIA Digital Archive - Demo

Controlled Vocabulary Maintenance Tool

Creating Video Segments and Annotations

- Contributor uses Annotator's Workbench software to segment and annotate video

- Starts with continuous video stream

- Divides stream into significant events

- Annotates (describes) each event

  - Free-form text descriptions

  - Controlled vocabulary fields

Video Segmentation Hierarchy

EVIA Digital Archive - Demo

The Annotator's Workbench

Future
Software Development

- Make applications available as open source

- Ability to add supplemental materials to annotator's collection

- Repurposing Video Segments

- Uncompressed video for archival storage

Future Software Development

- Integrate Workflow Management Tool and Technical Metadata Collection Tool

- New Features for Web User Interface

- FEDORA Integration through Indiana University Digital Library Program's Infrastructure Project