HOME | ABOUT D-LIB | CURRENT ISSUE | ARCHIVE | INDEXES | CALENDAR | AUTHOR GUIDELINES | SUBSCRIBE | CONTACT D-LIB

The Conference on World Affairs Archive Online: Digitization and Metadata for a Digital Audio Pilot

Michael Dulock and Holley Long
University of Colorado at Boulder
{michael.dulock,holley.long}@colorado.edu

Printer-friendly Version

Abstract

The University of Colorado Archives holds a substantial collection of audio recordings from the Conference on World Affairs (CWA), held annually at the University since 1948. Recordings of Conference sessions from the 1950s to the present comprise approximately 8,000 hours of material on reel-to-reel, cassette, and audio-only VHS tapes. In 2009 the Archives, along with other units of the CU Libraries and the offices of the CWA, began the pilot phase of a project to digitize these materials to make them accessible to the public. Between December 2009 and March 2010, the pilot produced 80 digitized recordings, with 15 receiving full metadata provision and presentation in the CU Digital Library in time for the 2010 Conference on 05 April. This paper describes the history and significance of the Conference and the collection; the project team; the planning and funding of the pilot; physical characteristics of the collection; the digitization rationale, specifications and process; metadata design and creation; and delivery of the content to the public.

Introduction

In 2009 the University of Colorado (CU) Archives, together with other units of the CU Libraries and the offices of the Conference on World Affairs (CWA), began a project to digitize a substantial collection of audio recordings from the Conference. This project not only increased access to a significant resource, but it also gave the team an opportunity to develop workflows for digitizing audio collections. While Libraries staff had extensive experience digitizing image and text based collections, audio digitization presented new challenges. This paper presents a case study of the CWA Digital Audio Archive with the intention of sharing the Libraries' experiences with other institutions that may be embarking on their first audio digitization project. Specifically, this paper establishes context for the Conference, describes the collection, and details the planning, funding, digitization, metadata, and delivery of the material.

1. Conference/Content

Every spring, luminaries from a broad range of disciplines and thousands of attendees are drawn to Boulder, Colorado to participate in the CWA, an event that is unique in both breadth of topical coverage and format. The first conference was an impromptu gathering organized by Howard Higman, CU professor of Sociology, in 1948 featuring Louis Dolivet, editor of *United Nations World* magazine. While the early days of the conference focused on international affairs, the subject matter quickly broadened to include science, technology, the arts, politics, spirituality, pop culture, ethics, the environment, and much more.

Roger Ebert, a long-time speaker, describes it as the "the Conference on Everything Conceivable." (Conference on World Affairs, n.d.).

The week-long conference is comprised of approximately 200 panel sessions, plenary speeches, and performances. Unlike many conferences in which specialists deliver prepared speeches, CWA speakers extemporaneously discuss and debate a wide range of topics. The dialogue is always lively and the perspectives fresh, because panels are composed of speakers from a variety of disciplines and spheres — science, the media, academia, the arts, and politics to name a few. The list of former conference speakers reads like a who's who of the last 60 years, among them Steven Allen, Joseph Biden, Norman Cousins, Roger Ebert, Buckminster Fuller, Molly Ivins, Henry Kissinger, Charles Krauthammer, Paul Krugman, Annie Liebowitz, George McGovern, Arthur Miller, Ralph Nader, Eleanor Roosevelt, Studs Terkel, Norman Thomas, and Ted Turner.

The conference is free and open to the public, attracting as many as 92,000 people in 2010. Attendees are drawn in large part by the opportunity to interact with the panelists. In fact, audience participation is a key to the distinctiveness of the event; the question and answer period for most sessions lasts 30-40 minutes. Recognizing the educational value of the event, local universities and high schools encourage their students to attend and the conference organizers ensure that students are given time to interact with the conference speakers.

2. Significance

Organizers had the foresight to record the conference sessions since the late 1950s, capturing and preserving this extraordinary event for future generations. The resulting CWA archive, housed in the CU Archives, contains more than 8,000 hours of audio as well as conference programs, presenter's biographies, photographs, and other materials related to the conference. Sessions selected for the first round of digitization illustrate this collection's significance. Examples include the 1961 *NATO as Nation* panel led by Henry Kissinger, a 1972 rap session between journalist Timothy Findley and Black Panther founder Huey Newton, and the 1980 *Women as Soldiers* panel. Unlike most public speeches, these sessions were unscripted. Speakers spoke candidly and in some cases found that their own opinions were forever changed by the debate. For instance, speaker Samuel T. Cohen, a nuclear weapons tactics theorist and Pentagon consultant, writes

> [S]ome 17 years later, the Conference hasn't changed, but I think I have. I have gone through a reappraisal of my views. I no longer believe in going out into the world and dropping bombs on Peking, Moscow. I'm still a nuclear Hawk when it comes to self-defense, but I've changed my mind about the containment of communism; I no longer believe in the domino theory. This metamorphosis came about because of the Conference (Higman, n.d., quoted in Montgomery, 2008).

The CWA archive not only captured the spontaneous and candid musings of some of 20th century's most influential figures, but it also has traced the evolution of certain issues over time. For example, nuclear proliferation, Israel and foreign policy in the Middle East, women's issues, the environment, and civil rights are examples of reoccurring themes discussed at the conference.

Thus, as an online resource, the CWA archive has great historic value and broad appeal to a wide range of individuals from a high school student learning about the ethics of science, to a movie enthusiast listening to insights from her favorite director, to humanities scholars doing research in any number of disciplines.

3. Team

The project team includes members from the CWA and several different departments within the University Libraries, as well as the digitization and web site vendors. The Director of Public Affairs at the CWA served as one of the project coordinators and provided not only extensive background on the collection and the conference, but also established lines of communication amongst team members from the different organizations. A member of the Committee for the CWA developed project goals and procedures, secured volunteer labor, solicited in-kind contributions from the digitization and website development vendors, and acted as the primary point of contact between the Libraries, the CWA and the vendors. The Faculty Director of the CU Archives served as the other project coordinator, contributing valuable insights as the collection manager. The Digital Initiatives Librarian from the Libraries Information Technology (LIT) department took part in discussions about the digitization of the analog audio, as well as worked closely with the Metadata Librarian from the Cataloging and Metadata Services department (CMS) to create a template for descriptive metadata and the presentation of the materials in the CU Digital Library (CUDL). Other members of LIT set up a server to stream the digital audio files. The Metadata Librarian selected the metadata schema for the descriptive metadata, built the template for metadata input, and consulted with the CWA regarding assembly of descriptive and technical metadata. A web design consultant developed a website to align the CWA archive with the CWA's main site's look and feel, and a high-end multimedia services firm converted the analog audio to digital format.

The partnership between Libraries and CWA personnel was close and heavily collaborative. Proximity afforded team members the opportunity to meet on short notice to discuss specifics throughout the project, and familiarity made communication fast and

easy. Questions and problems were swiftly addressed, and frequent messages among all parties kept everyone "in the loop."

4. Planning/Grants

The 2010 CWA archive pilot has its roots in earlier efforts to make this material available to the public in digital form. In 2008 and 2009, the University Libraries and the CWA unsuccessfully applied for a National Endowment for the Humanities grant to fund the digitization and online presentation of a small subset of the CWA audio recordings. Thereafter, the Libraries and the CWA looked for other funding sources. In October 2009, they met with a new project team to discuss the previous efforts to secure funding, review project specifics, and discuss potential new funding sources. The team considered options using local donations, hiring an outside project manager, and utilizing internal resources. However, the need for cost savings, as well as other factors, resulted in the use of a combination of local expertise and in-house labor to complete the project.

A sample set of 80 tapes already had been selected for the previous grant applications. This set served as the pilot material. Metadata formats were selected for all materials in question: Public Broadcasting Metadata Dictionary Project (PBCore) for the audio (which comprises the majority of the collection); qualified Dublin Core (DC) for text documents such as conference programs; and Visual Resources Association Core metadata 4.0 (VRA Core) for photographs and other images. CWA would provide some existing administrative and descriptive metadata and conference volunteers would provide additional item-level descriptive metadata for each audio session, for later conversion by CMS personnel into PBCore metadata within the Libraries' digital library metadata entry tool, Luna Inscribe. CMS staff would also have access to the digital audio streams in case further investigation was needed to provide suitable description. In addition, CMS would add a collection-level MARC record to the Libraries online catalog with links to the digital objects, providing yet another access point for users.

By December 2009, a plan was in place and work on the project could begin in earnest. The vendor selected for digitization, whose president has a long-standing interest in the Conference, generously agreed to digitize the eighty pilot tapes at near-cost using its facilities and equipment. In order to make some of the material available to the public online before the start of the 2010 Conference on April 5, all eighty tapes would be digitized, but full metadata provision and display in the CU Digital Library would only be completed for fifteen tapes, considered a reasonable subset to fully catalog and publicize before April.

Digitization was to begin in December and continue until all eighty tapes were done. Metadata template completion, allocation of storage space for the audio files, volunteer creation of item-level descriptions and metadata provision were to be completed from January to March. The target publication date was the last week in March, to allow some time for adjustments before the start of the conference.

5. Funding

Once the final plan involving the local digitization vendor was in place, funding for the cash costs of digitization was secured through a $50,000 donation to the libraries on behalf of the CWA initiative. Additional costs were mitigated by the digitization vendor, Wyndham Hannaway (President, GW Hannaway & Associates) and the web marketing vendor, John Cuddemi (President, smartmovebranding.com). Other work required of the CWA and the Libraries would be folded into existing workflows.

6. Physical Characteristics

In preparation for the project, the University Libraries' Media Specialist in the LIT department conducted an inspection of a sampling of the audio in the collection to ascertain media formats and assess the collection's condition. While the formats in the collection varied, conference organizers primarily used acetate-based reel-to-reel tape from the 1950s until 1964 when they switched to polyester-based reel-to-reel tape. Beginning in 1985, the conference began using audiocassettes to record the proceedings, and VHS from 1991 to 1994. These VHS contain both audio and video in some cases and audio only in other cases. Among the sampled media, the recordings lasted between 47 and 119 minutes, averaging 88 minutes per session.

The inspection invariably revealed issues with the physical condition of the sample. The majority of the reel-to-reel tapes surveyed had loose winds, uneven tape packs, popped strands, stretching and creasing, cinching, edge curl, and minor shedding, issues frequently caused by poor handling and playback techniques. While there was no indication of sticky shed syndrome — a condition in which the binder deteriorates — in the sampled reel-to-reel tapes, a slightly acrid odor emanating from several of the acetate-based reel-to-reel boxes indicated possible deterioration due to hydrolysis, a condition known as vinegar syndrome. The sampled audio cassettes were still in very good physical condition with the exception of a few mechanical problems with cases. The VHS tapes showed no discernible signs of physical damage.

The media specialist also listened to the selection of materials to assess the quality of the audio and rated sound quality on a scale of excellent, very good, good, fair, or poor. While none of the media sampled received a rating of excellent, the majority of the audio cassettes fell in the range of very good to good; the majority of the reel-to-reels were good to fair; and the

majority of the VHS recordings were fair to poor (Riberdy, 2007).

7. Digitization Rationale

Even though the collection survey suggested that the materials were in fairly good condition, the media have already exceeded their projected lifespan, and migration to a new format is necessary to preserve the content. In addition to the media's age, playback equipment for analog audio formats is becoming increasingly more difficult to acquire and maintain. Furthermore, this historically significant and unique archive was hidden due to the lack of publicly accessible surrogates. Even if researchers are able to discover the collection, they would have to travel to the CU Archives to listen to the materials. By digitizing the materials and providing online access, these barriers would be removed.

8. Specifications and Process

When the grant-planning phase began in 2006, the team reviewed existing literature on audio digitization guidelines and followed the recommendations outlined in the Collaborative Digitization Program's Digital Audio Best Practices (Collaborative Digitization Program [CDP] Digital Audio Working Group, 2006). This document was developed after a thorough investigation of existing guidelines and industry practices and provides recommendations for the entire digitization lifecycle from media preparation to quality control. The Libraries had used the CDP *Digital Audio Best Practices* for a previous project, so the team was experienced with these standards and best practices. CDP's *Digital Audio Best Practices* offered a three-tiered approach to its recommendations to account for the differences between audio collections and project goals. The team chose to digitize the CWA materials at the recommended 44.1 kHz and 24 bit, since these specifications adequately capture the frequency of spoken word.

Digitization services were donated by GW Hannaway & Associates. Reformatting the reel-to-reel and audio cassette tapes for this proof-of-concept was a straightforward process. Upon arrival at the studio, the technician inspected the media and determined that conservation treatment was not needed in order to proceed with digitization. He next played back the tapes on the appropriate equipment: a Braun TG-1000 C/4 for the open reel tape and a Nakamichi LX-5 cassette deck for cassettes. The technician then listened to the tape to establish the appropriate playback speed as well as to determine if there were any frequency loss and made adjustments as necessary. The tape head was then aligned for optimal playback. The audio was played through a Lucid 88192 converter to create the .aif master file. The audio technician used Soundtrack Pro 3 to optimize the audio and generate mp3 access files.

9. Metadata

Existing metadata

Because the CWA already had some metadata for the selected audio sessions to be used in the pilot, they sent the Metadata Librarian an Excel spreadsheet containing inventory details of the eighty sessions from which the fifteen pilot files would be selected. This document contained information for each session: tape number, tape type, date of the session, session title, moderator, speaker(s) and building location.

To supplement this data, the CWA enlisted its extensive network of volunteers to provide descriptive summaries and topical keywords for the pilot sessions. In addition, the volunteers provided time indexes with each description, providing timestamps and speaker changes as appropriate. Since the audio segments were not indexed by time but rather presented as single streaming files for each session, the timestamps provided additional value for users, since they could use the timestamps to locate a specific segment of interest.

Metadata format

The next step in metadata provision was to select a schema for presentation within the CUDL interface. The original grant proposal suggested the PBCore metadata schema for the audio portion of the collection. After examining the schema, the team determined that PBCore was indeed best suited to the collection and the Libraries. This schema includes descriptive and technical elements appropriate for both audio and video recordings, as the latter will be included in a future project phase. It is also relatively simple, while still providing a significant level of granularity. Most other CUDL collections feature rich metadata in a material-appropriate format, a practice which the team wished to continue with this project.

The PBCore schema (http://pbcore.org/PBCore/UserGuide.html) was adopted nearly intact, with a few departures from the standard necessitated by the needs of the collection and the Libraries. Since the CUDL presentation software does not display blank fields, the project team considered leaving all elements in the metadata template and simply leaving inapplicable ones blank. However, a shortened standard lacking unneeded elements would enable the catalogers who input the metadata to proceed more easily and rapidly. Thus this second approach was used. The Appendix contains the complete list of metadata

elements.

Intellectual Content

CWA metadata supplied a title for each session and a series title where applicable. To separate series title from session title for search purposes, series titles were recorded in a second title field, *titleSeries*. While "Series" is a valid *titleType* in PBCore, the team felt the "type" qualifier could confuse users. PBCore's *coverage* element has an accompanying element, *coverageType*, to distinguish between temporal metadata (session date) and spatial metadata (building location). However, the Libraries' presentation software counts each element as separate, and does not utilize dependent relationships such as *coverage* being modified by *coverageType*. This was not an issue for presentation purposes, but a search on a single *coverage* field in the CUDL would run against both temporal and spatial metadata. To resolve this issue, the Metadata Librarian replaced *coverage* with separate *coverage* fields, *coverageTemporal* and *coverageSpatial*, and eliminated *coverageType*.

The Metadata Librarian eliminated two additional elements in this group from the PBCore schema: *pbcoreAudienceLevel* and *pbcoreAudienceRating*. Archives and CWA personnel believed that the content was appropriate and appealing to a broad range of users, and thus felt that these elements were not useful.

Intellectual Property

After discussion regarding the nature of speaker participation in CWA sessions, the team decided to eliminate the elements *creator* and *creatorRole* while retaining *contributor* and *contributorRole*. This decision, while partly influenced by semantics, was also influenced by the controlled terms offered for both *role* elements in PBCore itself. Many PBCore elements draw from schema-specific pick lists, which contain many terms appropriate to the element. The terms available for the *creator* element, however, were not sufficiently specific. The desired terms for describing speakers in CWA sessions were "speaker" and "panelist". Both terms appear in the *contributorRole* pick list, but neither is available for *creatorRole*. No other terms for *creatorRole* fit the project's needs, so the team decided that the element for speakers would be *contributor*. The semantic factor was explained by CWA staff as part of the culture of the conference, which is based around discussion of a topic rather than presentation of prepared remarks. In addition, using *creator* alongside *contributor* could imply a false hierarchy in speakers' relative importance, which also conflicts with the spirit of the conference.

Instantiations

Elements in the *Instantiations* section of the standard were largely retained, even in the case of elements which pertained to video only, in anticipation of including video in later phases of the project. The team did eliminate the *pbcoreDateAvailable* group, which was not deemed appropriate for the material in the collection since the audio files being presented were not produced for commercial broadcast.

Input standards

Metadata was entered into PBCore elements using multiple input standards. Names were entered according to AACR2 (Anglo-American Cataloging Rules, 2nd edition) guidelines, using established forms from the National Authority File when available. Authority records were not created for names which did not have an authorized form. Dates and times were entered in the ISO 8601 format (YYYY-MM-DD and HH:MM:SS-07:00, respectively, where the "-07:00" in the latter indicates local time zone relative to Coordinated Universal Time [UTC]). Many terms, both descriptive and technical, were taken from PBCore pick lists, in which case the terms were entered exactly as found in the lists.

Metadata content

This section provides details on session-specific metadata contained in selected elements of the schema. Not all elements are discussed, such as those reserved for possible future use (*alternativeModes, extension*). A number of fields contained the same metadata in every record. Those fields were filled out with "constant data" when the metadata template was created to save time and effort during cataloging. Most of those fields, such as *publisher* and *formatEncoding*, are also omitted in this section since they contain metadata which requires no elaboration.

pbcoreTitle

The metadata contained in the Excel spreadsheet provided by CWA staff contained session titles, though there is some variation in how *title* was used. With two exceptions, every session among the eighty in the first project phase had a title. That title was used in the PBCore title element, with the *titleType* of "Program" from the PBCore pick list. Records for sessions which were also part of a series include that information in *titleSeries*. Plenary sessions were considered series for this purpose. The two sessions that did not have a distinct title were part of an ongoing series called "Rap Session".

pbcoreSubject

The descriptive notes provided by CWA volunteers served as the basis for subject analysis of the audio sessions. The Metadata Librarian used the summaries and keywords to create one or more Library of Congress subject headings (LCSH) for each session. The subject headings added were mostly topical in nature, but included names as well (e.g., a session on Fidel Castro). For topical headings, *subjectAuthority* contained the controlled term "lcsh". For names, "naf" for the National Authority File was used. These authority terms were drawn from the Library of Congress MARC code lists for sources: lcsh from the Subject Source Codes and naf from the Name and Title Authority Source Codes.

pbcoreDescription

After proofreading and any necessary edits, the descriptive paragraph supplied by the CWA volunteers was placed in the *description* element, with *descriptionType* as "Program", a term taken from the PBCore pick list. Time lines were also put in the *description* field, with the *descriptionType* of "Segment Sequence". The time lines consist of entries showing the time stamp in minutes and seconds, and a short description of the section.

pbcoreRelation

The *pbcoreRelation* group of elements was included to provide links between related sessions, such as installments of a series. The fifteen sessions in the pilot, selected in part to demonstrate the variety of Conference topics, do not yet utilize these fields, but once additional sessions are added to the collection, *relation* will provide rich information about related resources.

pbcoreCoverage

Coverage elements were split into two separate elements with their types embedded in the elements themselves. *coverageTemporal* contains the date of the session in ISO 8601 format. *coverageSpatial* records the campus building in which the session was held using the official campus nomenclature for the building.

pbcoreContributor

The team decided that contributor would suffice for all individuals and roles associated with the sessions. The *contributor* element contains names in inverted order, and in the authorized form according to the Name Authority Cooperative Program (NACO) of the Program for Cooperative Cataloging (PCC) when an authorized heading was available. *contributorType* was set according to the level of participation of the individual, as well as the nature of the session. Based on conversations with CWA staff, speakers would be listed as "Panelist" (from the PBCore pick list) for sessions featuring several speakers with equivalent billing, and "Speaker" for sessions primarily or exclusively featuring one speaker, as well as all plenary sessions. The team also decided that panel moderators would be mentioned at the beginning of the descriptive paragraph rather than listed as contributors, except in those cases where the moderator is also a speaker.

pbcoreInstantiation

The instantiation group of elements is composed largely of fields that could be considered technical metadata, since they provide information about the physical carrier and the digital access copy. Most of these fields were filled in as expected according to the standard, often with constant data. The few with variable data are mentioned briefly here.

dateCreated — records the date the audio session was digitized

dateIssued — records the date the collection was published in the CUDL

formatPhysical — records the original tape format using terms from the PBCore pick list

formatLocation — records the box and tape numbers as used by the CU Archives

alternativeModes — empty at this time, but kept for future use, such as to contain transcripts

pbCoreDateAvailable fields were discarded, considered to be unimportant for this project since the content was never broadcast. *pbcoreFormatID* fields were used to record the direct URL for the session (*formatIdentifier*) and identify the CWA and CU Archives as the responsible parties (*formatIdentifierSource*).

Technical content

Expectations related to and recording of technical metadata have been something of an issue throughout the project. Originally the team hoped to be able to use Audio Engineering Society Core Audio Metadata (AES-X098-B) to record technical metadata for both archival and access files, but the standard is still under development and is not available to the public (Riley, 2008). As of this writing, technical metadata has only been captured in the Instantiations section of the PBCore documentation, which does record a significant amount of technical information (bit rate, sampling rate, file size, etc.). Some of this information is also

captured in the header of the archival audio file itself. As the project proceeds, the team hopes to work with the vendor to capture and store more granular technical metadata for the audio streams. Technical metadata for archival files will likely be stored within the *Extensions* element group.

10. Delivery

Users can access the CWA Audio Archive from two different platforms: the [CU Digital Library](#) or the [CWA website](#). The University of Colorado Digital Library has housed all collections digitized by University Libraries since its inception in 2005. As the central point of storage for digitized resources, users can search across collections to find a variety of materials on any given subject. For example, a user who searches "Africa" will not only retrieve a 1959 CWA plenary session delivered by the Kenyan Secretary General, but also the David Rumsey collection of African Maps.

The version of the software on which the CU Digital Library runs can provide "progressive download" for audio content, but the CWA organizers preferred streaming delivery for greater access control. Therefore, the audio is streamed from the Libraries' Flash streaming server. Users access the audio file via a link in the CU Digital Library metadata.

In addition to the Libraries' access point, the CWA organizers hired a design company to develop a Flash website with the CWA branding. This site not only connects the CWA archive to the CWA main site's look and feel, but also provided a fund raising medium to finance ongoing digitization.

11. Conclusion

The CU Archives and CWA are pleased with the pilot collection, though with almost 8,000 hours of audio left to digitize, work will continue for some time. Since the 2010 Conference, the digitization vendor has converted another one hundred hours of additional audio, with plans to digitize additional tapes as funding allows. Libraries' CMS will fold metadata production into the existing workflow using procedures created for the pilot as new audio files and descriptive summaries are created.

Now that the proof-of-concept collection has been completed, the team is considering ways to add functionality to the digital collection to create a more participatory experience for users. For example, the team is investigating ways to allow end-users to mark a specific passage in the audio and email it as a link to friends. The team also plans to add comments functionality, thus allowing users to discuss sessions online. These features will extend the participatory nature of the conference into the digital realm.

The partnerships between the Libraries' Archives, LIT and CMS departments, the CWA, and a generous and enthusiastic local vendor have enabled access to a rich resource which offers a broad, candid picture of the culture, politics, and current affairs of the last six decades. The University of Colorado and the CWA are committed to sharing as much of the CWA archive as possible with the wider public and scholarly communities.

References

[1] Collaborative Digitization Program Digital Audio Working Group. (2006). *Digital Audio Best Practices Version 2.1*. Retrieved August 31, 2010 from http://www.bcr.org/dps/cdp/best/digital-audio-bp.pdf.

[2] Conference on World Affairs. (n.d.). *63rd Annual Conference on World Affairs*. Retrieved from http://www.colorado.edu/cwa/information.html.

[3] Higman, Howard (n.d.). *In the Beginning: the History of CWA*. Retrieved from http://howardhigman.com/.

[4] Montgomery, Bruce. (2008). 40 Years of World Affairs; the Audio History of the Conference on World Affairs. Unpublished grant proposal to the National Endowment to the Humanities.

[5] Riberdy, M. (2007). *Technical analysis for the CWA pilot audio survey*. Unpublished technical report, University of Colorado at Boulder.

[6] Riley, J. (2008). *Metadata for Audiovisual Materials and its Role in Digital Projects* [PowerPoint slides]. Retrieved from http://www.olacinc.org/drupal/conference/2008/WorkshopDescriptions.html.

Appendix: Metadata Specifications

(Exact content in quotation marks.)

INTELLECTUAL CONTENT	
pbcoreIdentifier	
identifier	Access file name
identifierSource	"CWA/Archives, University of Colorado at Boulder"
pbcoreTitle	
title	Session title
titleType	"Program"
titleSeries	Series title, if applicable
pbcoreSubject	
subject	Subject heading
subjectAuthorityUsed	"lcsh" for topics, "naf" for names
pbcoreDescription	
description	Descriptive paragraph/segment time line
descriptionType	Description type
pbcoreGenre	
genre	Reserved for future use
genreAuthorityUsed	Reserved for future use
pbcoreRelation	
relationType	Reserved for future use
relationIdentifier	Reserved for future use
pbcoreCoverage	
coverageTemporal	Session date
coverageSpatial	Session building location
INTELLECTUAL PROPERTY	
pbcoreContributor	
contributor	Contributor name in inverted order
contributorRole	Contributor role
pbcorePublisher	
publisher	"CWA/Archives, University of Colorado at Boulder"
publisherRole	"Publisher"
pbcoreRightsSummary	
rightsSummary	"All rights reserved"
INSTANTIATIONS	
pbcoreInstantiation	
dateCreated	Digitization date in ISO 8601 format

dateIssued	Digital file publication date in CUDL
formatPhysical	Format of original physical carrier
formatDigital	"audio/mpeg"
formatLocation	CU Archives box & tape number
formatMediaType	"Sound"
formatGenerations	"Audio/Original recording"
formatStandard	"MPEG Audio"
formatEncoding	"mp3"
formatFileSize	File size in bytes
formatTimeStart	Session start time in ISO 8601 format
formatDuration	Duration of each session in ISO 8601 format
formatDataRate	"128 kbps"
formatBitDepth	"16 bits"
formatSamplingRate	"44.1 kHz"
formatFrameSize	Reserved for future use
formatAspectRatio	Reserved for future use
formatFrameRate	Reserved for future use
formatColors	Reserved for future use
formatTracks	"1 audio track"
formatChannelConfiguration	"dual-channel mono"
language	"English" for all sessions to date
alternativeModes	Reserved for future use
pbcoreFormatID	
formatIdentifier	Direct URL for session
formatIdentifierSource	"CWA/Archives, University of Colorado at Boulder"
pbcoreAnnotation	
annotation	Reserved for future use
formatIdentifierSource	"CWA/Archives, University of Colorado at Boulder"
EXTENSIONS	
pbcoreExtension	
extension	Reserved for future use
extensionAuthorityUsed	Reserved for future use

About the Authors

Michael Dulock is Metadata Librarian and Assistant Professor at the University of Colorado at Boulder. His research focuses on metadata design & creation for digital collections and metadata schema selection. Prof. Dulock has also published in the Journal of Library Metadata and Cataloging & Classification Quarterly. He received his M.L.S. from Drexel University in 2006 and also holds M.A. and B.A. degrees in Classics from the University of Colorado at Boulder.

Holley Long is the Digital Initiatives Librarian and Assistant Professor at the University of Colorado at Boulder. Her research interests include usability and user centered design of digital collections and services.