

CASE 12

Podcasts in the Archives: Archiving Podcasting Content at the University of Michigan

AUTHOR: **ALEXIS A. ANTRACOLI**
University of Michigan

PAPER DATE: JULY 2010

CASE STUDY DATE: 2010-ongoing

ISSUE: This case study examines the challenges involved in developing best practices and workflows for archiving and preserving podcasting content created by the University of Michigan. One major issue involved establishing standards of practice for ingest, storage, and access, especially the generation and storage of appropriate descriptive, technical, and preservation metadata. Another challenge centered around developing the necessary technological infrastructure to support an Open Archives Information System (OAIS)-compliant system.

KEYWORDS: Podcasting content, Born-digital audio, Born-digital video, UARP, OAIS-compliant process, Preservation, Metadata

Copyright by Alexis A. Antracoli

Institutional Context

The University of Michigan was established in Detroit in 1817, and relocated to Ann Arbor in 1837. In addition to the main Ann Arbor campus, the University of Michigan has two satellite campuses. The University of Michigan-Flint opened in 1956, and the University of Michigan-Dearborn followed in 1958. In 2007, the total university enrollment for all three campuses approximated 55,000 undergraduate and graduate students. The University is comprised of 19 schools and colleges. In 2010 over 13,000 degrees were awarded. The university's research expenditures total over \$700,000,000.

The University Library contains 29 libraries within its system, including the Bentley Historical Library. Established in 1935 by the University of Michigan Regents, the Bentley Historical Library has two functions: to serve as the official archives of the University and to document the history of the state of Michigan and the activities of its people, organizations and voluntary associations. The Bentley is comprised of three divisions: the Michigan Historical Collections, the University Archives and Records Program (UARP), and Access and Reference Services. The Bentley Library reports to the Provost and Executive Vice President for Academic Affairs.¹ Bentley Library staff members who were involved at some level in this case study included Francis Blouin, Nancy Bartlett, Nancy Deromedi, Greg Kinney, Len Coombs, Brian Williams, and Olga Virakhovskaya. Jim Ottaviani and Shane Beers of the University Libraries were also instrumental in guiding the project. Alexis A. Antracoli served as the digital preservation intern on the project, and Caitlin Holman provided the technical expertise needed to develop the metadata extractor.²

Issue

This case study examines the challenges involved in developing best practices and workflows for archiving and preserving podcasting content created by the University of Michigan. One major issue involved establishing standards of practice for ingest, storage, and access, especially the generation and storage of appropriate descriptive, technical, and preservation metadata, and developing the necessary technological infrastructure to support an OAIS-compliant system.

Background: History of Podcasting at the University of Michigan

Technological and communications innovation has historically been a hallmark characteristic of the University of Michigan. The first telephone was installed on campus in 1881, and in 1906 the Regents approved installation of a campus telephone exchange. By 1939, the University was home to more than 1400 telephones. Only twelve years later, in 1951, researchers developed the University's first computer, the MIchigan

¹ Thanks to Nancy Deromedi for this overview of the institutional background of the University of Michigan and Bentley Historical Library.

² This project would not have been possible without the collaboration of the offices of Michigan Marketing Design and Michigan News and Information.

Digital Automatic Computer (MIDAC). By the end of the 1950s the university had established a computing center. Founded in 1950, [University of Michigan Television Center](#) (TVC), later Michigan Media, distributed programming via educational and commercial television stations and established the University as one of the first in the country to produce and distribute educational television programs.³ Email came to campus in 1981 via the CONFER system. Over the past three decades, the university has established, a campus computing program for students, and re-chartered of the School of Library and Information Science as the School of Information (1996). The history of podcasting thus constitutes yet another chapter in Michigan's history of technological innovation in communications and information dissemination.⁴

Podcasting began at the University of Michigan in the fall of 2004, when Jared Van Ittersum, a first-year dental student, requested that all Dental School lectures be recorded and made available to students for study purposes. Dr. Lynn Johnson, Assistant Dean for Informatics and Innovation at the Dental School, approved a pilot program to investigate the benefits and costs of video-recording dental school lectures. After discovering that audio recordings were more valuable to students and determining that students themselves could undertake most of the actual recording, the Dental School launched a program to begin recording classroom lectures for study purposes. The Dental School's efforts were so successful that the institution was invited to contribute to the initial launch of Apple, Inc.'s iTunesU project. The University of Michigan School of Dentistry accepted the invitation and joined with the University of Missouri, University of Wisconsin, Duke, Brown, and Stanford, to help launch Apple's iTunesU in 2005. While the original Dental School lectures were intended to be restricted to student use, in 2008 the University of Michigan began releasing public content through iTunesU. In 2009 the university also began releasing content through a University of Michigan YouTube channel.⁵

By 2010, the University of Michigan's iTunesU presence consisted of content provided by dozens of offices and departments across the University in fields ranging from English literature to physics. Additionally, several schools and offices within the University, including the University of Michigan News and Information Service, the Ross School of Business, and the Open Michigan Initiative decided to sponsor their own YouTube channels, which supplement content released on the main University of Michigan channel.⁶

³ See Media Resources Center (University of Michigan) records, 1948-1987 at the Bentley Historical Library.

⁴ See Nancy Barlett, et al. eds., "A Century of Connectivity at the University of Michigan," *Bentley Historical Library Bulletin*, 55 (December 2007).

⁵ This history is based on information available at the University of Michigan Dental School's website (<http://www.dent.umich.edu/dentalinformatics/services/videotaping>), the University of Michigan and Dental School YouTube channels (<http://www.youtube.com/user/um> and <http://www.youtube.com/user/UMichDent>), and the Dental School Podcast "A Brief History of Podcasting at the University of Michigan School of Dentistry," which is available through iTunesU.

⁶ More information on the type of content released by the university through these two mediums can be found by visiting the University of Michigan's iTunesU site and the university's YouTube channel.

Podcasting Timeline

- 2004** Podcasting begins at the School of Dentistry.
- 2005** University of Michigan creates first YouTube channel.
- 2005** School of Dentistry joins with iTunesU.
- 2008** University of Michigan starts releasing content to iTunesU.
- 2009** School of Dentistry begins releasing videos via YouTube.⁷

Challenges

Born-digital audio and video that has been created and disseminated as podcasts present the university archives with several challenges. Significant issues include the need to define the podcasting genre, to understand the special handling needs of born-digital audio and video files, to establish procedures for ingest, and to develop tools for generating descriptive, technical, and preservation metadata.

Methodology

The University Archives and Records Program (UARP) of the Bentley Historical Library created an internship through Elizabeth Yakel and Magia Krause's graduate Digital Preservation course. The internship provided an opportunity for a University of Michigan School of Information student to research podcasting preservation challenges and to develop a plan for the archiving and preservation of podcasting content. This internship, amounting to 90 hours, occurred in the winter term in 2010. In collaboration with the staff of UARP, including the Head of Division and the Digital Archivist, the intern researched existing literature and projects on the topic and created an appraisal and preservation plan.

Defining Podcasting

The term "podcast" is loosely connected with digital audio and video. The staff of UARP, however, focused on identifying a working definition for the purposes of our current project. Research indicated that podcasts have been defined most commonly as a form of syndicated, web-based audio or video content. They usually are distributed via web syndication such as an RSS feed.

Additional University of Michigan sponsored YouTube channels can be identified by browsing Subscriptions.

⁷ The timeline was created from the same sources at the preceding history.

Existing Efforts

A survey of the existing efforts to use and archive podcasts revealed that, while archival institutions are beginning to use podcasting as a form of outreach, no institution has yet undertaken a systematic effort to create a plan for the appraisal, ingest, retention, and dissemination of podcasting content.⁸ This review of peer institutions included all of the Big Ten universities, as well as Harvard University, Cornell University, and the University of California. Other non-academic institutions included the Internet Archive, the Library of Congress, the Museum of Broadcast Communications, the National Archives and Records Administration, and WGBH public television. Professional and scholarly journals also revealed a lack of published research results on podcasting.

The results of the initial survey determined that many institutions, including the Library of Congress, are using podcasts as outreach tools, but that only three archives actually archive podcasts. These institutions were the University of Iowa, Cornell University, and the Internet Archive. At Cornell University, the University's institutional repository holds ten podcasts on varying topics, but most are related to library outreach.⁹ The University of Iowa's efforts are more systematic, but on a small scale. The Iowa Digital library currently holds twelve podcasts, ten of which are related to the Iowa City Summer Writing festival. Efforts to record author events at the festival grew out of a collaborative effort between Iowa's International Writing Program and School of Library and Information Sciences' "Virtual Writing University Archive." The podcasts can be accessed at the "Virtual Writing University Archive" or the [Iowa Digital Library](http://ecommons.cornell.edu). The history of the project indicates that Iowa's archival podcasts grew out of the "Virtual Writing University Archive," rather than as part of a concerted effort to archive podcasts as a genre.¹⁰

The most comprehensive effort to archive podcasts is the Internet Archive, which accepts digital audio and video content from any registered user. Although it holds the most diverse array of podcasting content of any digital archive, the Internet Archive lacks a sustained focus associated with institutional archives.

Appraisal

Our appraisal process combined institutional functional analysis with an assessment of the University and Archives and Records Program's collecting strengths and weaknesses, as well as an assessment of the popularity of individual podcasts. In particular, Helen Samuels' approach to documenting the functions of a university provided a framework for assessing the podcasts. Because most of the podcasting content focuses on classroom lectures, outreach, and student recruitment, the content falls mostly under the following functions: confer credentials, convey knowledge, sustain the institution, and provide public service. In assessing the strengths and weakness of the UARP collection, it

⁸ See J. Gordon Daines and Corey L. Niemer, eds., "The Interactive Archivist: Case Studies in Using Web 2.0 to Improve the Archival Experience," <http://www.lib.byu.edu/sites/interactivearchivist/>.

⁹ <http://ecommons.cornell.edu>

¹⁰ <http://digital.lib.uiowa.edu/vwu/>

became clear that UARP is strong in documenting technological innovation, as well in acquiring information from the offices of News and Information and the President. The records produced by these offices generally document the process of sustaining the institution and providing public service. UARP is weaker in documenting classroom instruction, foreign language outreach, and some of the smaller schools within the University, such as the Ford School of Public Policy. Documenting classroom teaching has always been difficult, and is often best done primarily through acquiring faculty papers and departmental records about curriculum. Podcasts represent a new way to document this function. In determining offices to initially approach for podcasting content, these considerations were combined with assessments of the popularity of various podcasts on iTunes and YouTube.

With these considerations in mind, beta-testing began with a choice of two YouTube videos and three iTunes selections (two audio and one video). These two means of dissemination afford slightly different content. Thus, it was necessary to consider both of them in selecting content for targeted ingest. The selections were based on content that was deemed valuable according to the appraisal standards described above. For example, foreign languages remain under-documented in the University of Michigan archives, and the UM en Español video chosen was the second most viewed on the UM en Español YouTube channel. It also documents University outreach efforts to Spanish-speaking prospective students. On the other hand, “University of Michigan-iPhone as a Musical Instrument” podcast serves to document the history of innovation at the University, a strength of the collection, and it received 81,196 views on YouTube as of May 25, 2010. The iTunes selections included two audio recordings, one from the Ford School Citigroup Foundation Lecture Series and one that is a lecture by the popular and award-winning English professor Ralph Williams. Overall, the selections represent variety in length and type of content, while also representing the content that UARP will likely want to continue to ingest and preserve over the long term.

UARP contacted the offices of Marketing Design, News and Information, the President’s Office, the Ford School of Public Policy, and the English Department to discuss practices for archiving podcasts. The first three units build on UARP’s existing strengths. Marketing Design also produces the UM en Español materials whose addition will strengthen a weak area within the collection. The Ford School represents a smaller and less documented school within the University, while the English Department podcasts will assist in documenting classroom teaching and curriculum.

Processing Overview

After making decisions about the podcasts to target for ingest, the staff of UARP worked on creating an OAIS-compliant process for handling podcasting content. Ingest, preservation, and access are generally discussed as separate concerns, but this project demonstrated that they are intertwined, and that one must consider both preservation and access needs in developing an ingest process for digital objects.

Ingest

The ingest process will include the transfer of select podcasts from the offices of origin to UARP, including a copy of the digital object and associated metadata. At this time, UARP is working to build a web form that will allow offices to upload batches of digital objects. The details of the ingest process, including file format and metadata standards, were shaped by long-term preservation needs. The offices of origin will upload their objects in approved file formats with a minimum set of associated descriptive, technical, and preservation metadata. This will allow UARP to guarantee some level of preservation over the long term. All metadata will be embedded in the file header for extraction via the web form.

Preservation: File Formats, Metadata, and Storage

In order to ensure that the ingested files would survive over the long term, UARP needed to determine what file formats and metadata were best for this purpose. Extensive research into file formats included a search of the professional literature, discussions with government agencies, and major universities that have digital preservation programs, and consultation with Shane Beers, the University's Digital Preservation Librarian. We concluded that .aiff or .wav files are recommended for audio files and uncompressed .mov, .avi, or Motion JPEG 2000 are best for video, although video standards are still developing. If files could not be obtained in these high-quality formats, UARP decided to ingest .mpeg and MPEG2 or MPEG4 files wrapped in .avi or .mov. (See Appendix A).

Metadata standards were taken from a combination of resources in an effort to ensure that the descriptive, technical, and preservation metadata necessary to maintain and render the files in the future would be maintained. In particular, UARP will combine aspects of the PREMIS data dictionary, Dublin Core descriptive metadata, and a custom set of technical metadata for long-term archival storage. (See Appendices A and B). Two preservation copies of the ingested files will be kept by UARP on University server space in separate locations. Metadata will be stored separately from the digital objects in UARP's BEAL database.¹¹

Access

Access copies of the files will be provided through [Deep Blue](#), the University of Michigan's institutional repository. Each discrete item will have a set of Dublin Core metadata containing the following fields: creator, contributors, series, title, date of creation, description, rights, format, type, and language. The object identifier, bitstream url, and extent (size) are automatically logged and tracked by Deep Blue upon upload. (See Appendix C). Links to the finding aid will be provided at the collection level in Deep Blue, and EAD finding aids available through the Bentley website will contain links that take patrons directly to the specified item. Additionally, we will work on providing links to Deep Blue from the University Libraries' catalog, MIRLYN.

¹¹ BEAL is UARP's accession tracking system, location guide, and donor database.

Workflow

In order to process these files from ingest to access, members of the UARP staff and eventually student workers, will need to follow a workflow that meets the above requirements. Many of the steps in the process, however, will be automated.

Once files have been uploaded to the web based metadata extractor, descriptive and technical metadata will be extracted from them and exported to an XML file. A member of the UARP staff will then import the XML files into the BEAL database. Once the XML file with all relevant technical and descriptive metadata is imported into BEAL, the digital object itself will be validated using JHOVE if a module exists for the file type. A checksum will also be performed. Associated preservation metadata, including events, agents, fixity, and content location will be entered into BEAL by hand.

After the initial ingest, validation, and creation of metadata, a member of the UARP staff will copy and reformat, when necessary, the digital objects. UARP will create one additional copy of a preservation-worthy object, as well as a lower-quality access copy. Both a high quality and lower-quality copy will go into Deep Blue. Meanwhile, two preservation copies will be stored in different locations on UARP servers, and the access copy will be uploaded to Deep Blue with associated Dublin Core metadata. Deep Blue also keeps a backup copy of the access version.

Partners

The staff of UARP worked closely with several individuals from the University libraries, as well as various content creators. Developing an implementable process for handling podcasts required collaboration with the head of Deep Blue, the University's institutional repository, the University library's Digital Preservation Librarian, and the offices of Marketing Design and News and Information. The head of Deep Blue was able to customize the deposit form to match UARP's metadata needs, while the University's Digital Preservation Librarian provided helpful advice on current file formats and metadata standards for preservation purposes. Currently, UARP is collaborating with Marketing Design and News and Information in order to ensure timely transfer of designated podcasting content to the archives, and to develop a standard process by which those transfers can occur. The project would not have been possible without the cooperation of these partners and library staff.

Resources Required

In order to develop access, preservation, and ingest procedures and standards, UARP set up an internship through the Institute of Museum and Libraries Services/University of Michigan project "[Engaging Communities: Fostering Internships for Digital Preservation and Curation](#)." The intern, a University of Michigan School of Information student, spent approximately 90 hours researching digital audio and video preservation, appraising content, and developing standards and a proposed workflow for the project. Additionally, another School of Information student, employed by the Bentley Library, is currently

creating the customized web form and metadata extractor necessary to implement the system.

Analysis

Deciding how to handle podcasting content generated several challenges for the University Archives and Records Program. Developing a set of file format and metadata standards was particularly challenging because standards for these formats, especially video, are less well-developed than for other types of digital files. This challenge required a significant time investment into researching emerging standards for preservation and access. However, after exploring the professional literature and practices of other institutions, UARP was able to develop standards that made sense for its purposes.

Another major challenge was determining what technological infrastructure would be necessary to implement an ingest procedure. Ultimately, some collaboration with other units was necessary. Fortunately, UARP also had someone on staff who could develop a customized, web-based metadata extractor to ingest a wide variety of digital objects. This has helped to automate the metadata creation and ingest process. Metadata is created by offices of origin, and will be extracted automatically through a web form before ingest into the BEAL database.

Finally, cooperation with the offices of origin was necessary in order to ensure that the ingest process could be implemented. Currently, UARP is working with the University of Michigan's Office of News and Information to develop a customized ingest process. The archives will continue to work with other offices to ingest podcasts through the web-based metadata extractor. Creating productive alliances with other campus offices can be a challenge, but UARP has found the offices of origin to be interested and cooperative.

Without the cooperation of the creating and disseminating offices, and the resources to develop a technological infrastructure for ingest, this project would not have been possible. Currently, UARP is planning to ingest its first batch of approximately 300 podcasts from Michigan News and Information. In the summer of 2010, this beta ingest will form the basis for determining a workflow for processing podcasts from the point of ingest to the point where they are accessible to researchers, including plans for storage and preservation. This process will then form the basis for creating a workflow for processing digital objects and collections within the University Archives and Records Program.

Appendix A

Audio and Video Podcast File Formats and Technical Metadata

Audio Ideal

Format: Broadcast wave format (.wav)

Metadata (to be included in file header):

Creator
Contributors
Title
Date of Creation
Description
Format
Extent/Size
Creating Application
Duration
Bit Depth
Sample Rate
Number of Channels
Audio Data Orientation
Sound Channel Map
 Sound Field
 Sound Channel Assignment
 Channel Number
 Sound Map Location.

Audio Minimum

Format: mp3 format

Metadata (to be included in file header):

Creator
Contributors
Title
Date of Creation
Description
Format
Extent/Size
Creating Application
Duration
Bit Depth
Sample Rate
Channel (number and information)
Sound Field

Video Ideal

Format: Motion JPEG 2000 or uncompressed .mov or .avi in MPEG-2 format.

Metadata (to be included in file header):

Creator
Contributors
Title
Date of Creation
Description
Format
Extent/Size
Creating Application
Duration
Bit depth
Channel (number and information)
Sound Field
Color
Compression
Data rate
Frame
Resolution

Video Minimum

Format: MPEG 4 (.mp4) in .mov or .avi wrapper (compressed .mov or .avi files that are not encoded in the MPEG 4 format are not acceptable)

Metadata (to be included in file header):

Creator
Contributors
Title
Date of Creation
Description
Format
Extent/Size
Creating Application
Duration
Bit depth
Channel (number and information)
Sound field
Frame
Resolution

Appendix B

PREMIS Data Dictionary: Elements to Track

The following elements should be tracked using the PREMIS data dictionary:

Object Entity

- Object Identifier
- Preservation Level
- Fixity
- Creating Application
 - Name
 - Version
 - Date Created
- Storage
 - Content Location

Event Entity

- Event Identifier
- Event Type
- Event Date and Time
- Event Detail
- Event Outcome Information (for validation)
 - Event Outcome
- Linking Agent Identifier
 - Type
 - Value
 - Role

Within the event entity the following events should be tracked using the Event Entity of the PREMIS Data Dictionary:

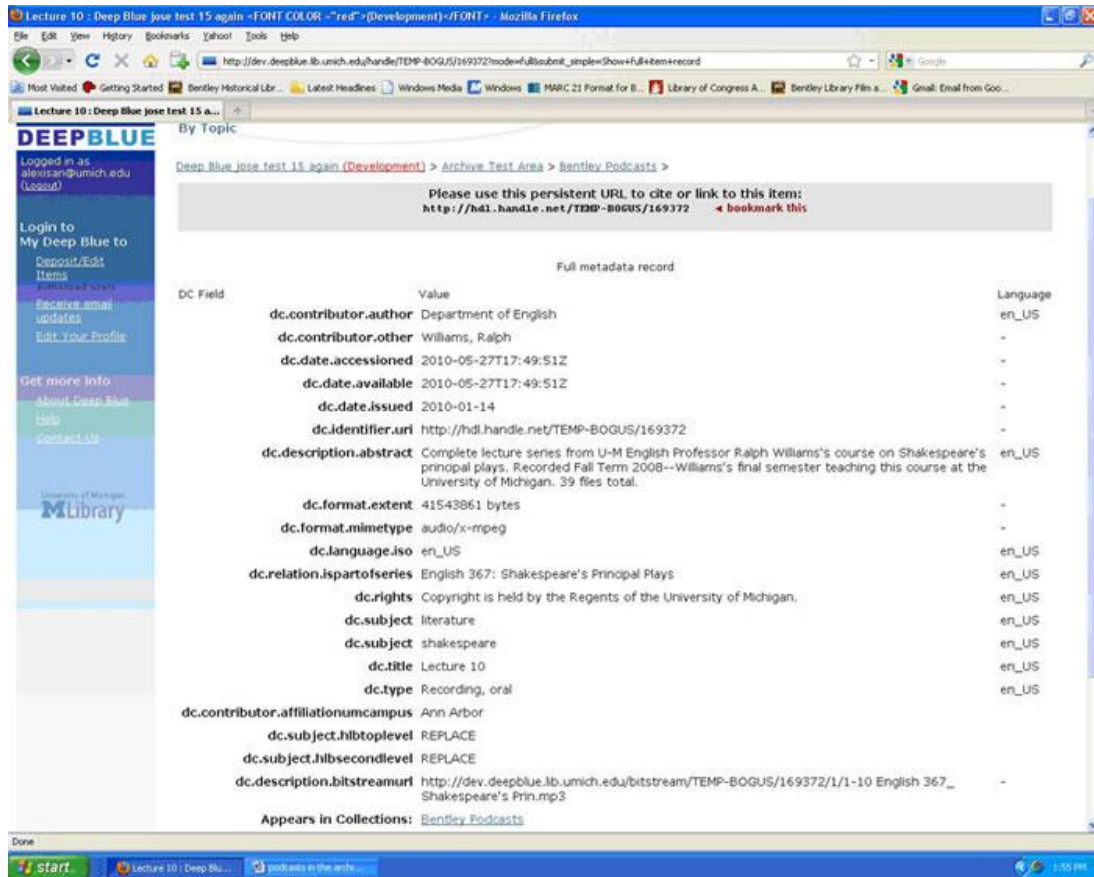
- Capture
- Validation (if done)
- Message digest calculation
- Fixity Check
- Ingest
- Migration

Agent Entity

- Agent Identifier
 - Type
 - Value

Appendix C

Dublin Core Descriptive Metadata (as it will appear in Deep Blue)



V. Keywords (optional)

The following keywords are suggested, in part for ease of searching as Campus Case Studies accumulate. Select all that apply.

- Access interface
- Administrative information systems
- Administrative records
- Appraisal issues**
- Authority/Mandate
- Blogs
- Content management systems
- Course management systems
- Custodial issues
- Data format issues

Data integrity issues
Data longevity issues
Disaster planning
Digital photographs
Digital moving images
Document imaging systems
Email
Faculty papers
Faculty research data
File format issues
Implementation planning
Institutional repository
Legacy systems and media
Metadata
Podcasts
Policy documents
Publications (catalogs, magazines, reports, etc.)
Recordkeeping systems
Resource issues (monetary, etc.)
Social networking systems
Standards
Student-created records
Student records
Vital records
Websites
Other (enter keywords) _____

Thank You!

Does your university archives have born-digital records?

Share how you are effectively managing these digital records by submitting a case study to Campus Case Studies.

Visit www.archivists.org/publications/epubs/CampusCaseStudies/.