DIY DIGITAL RADIO PRESERVATION:
AN INTRODUCTION TO CITR'S AUDIO ARCHIVING PROJECT

Shyla Seller, University of British Columbia, Canada

1.   Introduction

In 1937, the University of British Columbia's Alma Mater Society inaugurated Varsity Time, a weekly half-hour radio program on CJOR, meant to provide information about campus activities to residents of Vancouver and BC. From this program grew the Radio Society collective, which aimed to provide a training ground for journalists and broadcasted a weekly news program "UBC Digest" across the country in the 1950s. The society became known as CYVR — UBC Radio in 1969, when it began broadcasting from the Student Union Building. In the mid-seventies the station was reopened after a Canadian Radio-television and Telecommunications Commission (CRTC) shutdown as CiTR, which continues to broadcast today in FM at 101.9.

CiTR's collection of over six hundred reel-to-reel audio tapes is unique and diverse, consisting of radio programming including but not limited to: live performances, interviews, sports and news reports, radio plays, lectures, regular radio shows, public service announcements, and station advertisements. They have many recordings of musical performances by local bands going back to the early 1980s that have significant cultural value and are not found anywhere else. CiTR has never been overly concerned with sound quality, and has operated on a shoestring budget doing as much as they can with limited resources. Approximately three hundred of their tapes are housed at the University Archives at UBC, and another three hundred are in the offices of CiTR, in the Student Union Building, awaiting transfer to those archives. CiTR has a volunteer with recording engineering experience digitizing reels to preserve and provide access to their collection of audio material. At the same time, they are making current, born-digital programming available as mp3s for streaming or podcast from their website. This paper will review current issues and best practices in digital audio preservation, introduce the station's current production of both digitized and born-digital audio content, and provide recommendations to CiTR for incorporating best practices into their existing workflow to optimize the long-term digital preservation of the records of their programming activity.

The station is primarily run by volunteers and has limited resources, financial or staff, to contribute to a digital preservation program. Therefore the perspective and analysis in this paper will be tailored to a small, bare-bones, do-it-yourself organization, keeping in mind the limitations of CiTR's working environment and attempting to present ideas that can conceivably be incorporated into the existing climate of the non-profit organization. My approach is informed by case study research conducted in InterPARES 2, which, in reference to digital preservation in the arts, found that "standardized preservation criteria and procedures may not be effective or appropriate and that preservation decisions may, in many instances, need to be addressed on a case-by-case basis."[64] I use the definition of digital preservation from Luciana Duranti, i.e., the specific process of maintaining digital materials during and across different generations of technology over time, irrespective where they reside.[65]

2.   Issues in Digital Audio Preservation

Samuel Brylawski reminds us "Sound has been recorded digitally since the 1970s, when pulse code modulation (PCM) became an accepted method of recording by audio engineers and producers."[66] Digital audio files are created and distributed in a myriad of ways, existing in disparate formats, qualities, on CDs, DATs, as files on phones, portable music players, or computers. In fact, Brent Lee writes that the single thing held in common by digital audio files is that "they all contain a stream of numbers that represent changes in the amplitude of sound pres-

---

64   Hackett et al., 23.
65   Duranti, "The Preservation of Digital Records".
66   Brylawski, 53.

sure over time."[67] File-based born-digital recording is the dominant method of contemporary audio production, while the fragility and threat of obsolescence of machine-dependent historic audio formats, as well as the chemical instability of materials composing CDs and DVDs, has made digital reformatting of audio records a standard tool of preservation and conservation.[68]

However, audio digitization and preservation projects are challenging for organizations of any size to undertake. They require new knowledge and training, ongoing funding, dedicated staff, extensive IT support, project management planning and workflows, quality control, attention to standards, appraisal, and arrangement and description, as well as ongoing accessible long-term storage. This list of requirements is not exhaustive. There are best practices and guidelines for workflow, such as those developed at Indiana University's Sound Directions and IASA, and conceptual models for digital preservation like OAIS and InterPARES' Chain of Preservation (COP), but typically these presume an existing record-keeping system is in place, and are prepared with resources from larger library and archival institutions in mind. They would not easily be implemented by a volunteer-based radio station like CiTR. Even within organizations with dedicated programs, budgets, and staff in place, the lack of knowledge about the recording and manipulation of digital audio files, the ear training and technical experience necessary to discern audio quality and faithfulness to originals, the threat of proprietary file formats not being readable in the future, the lack of standards, accuracy, and completeness for files' metadata, inadequate appraisal and description, appropriate and affordable storage repositories, and the sheer expense of the undertaking make long term preservation a daunting prospect for any organization with significant audio holdings.

Technological and organizational issues aside, there are also issues associated with the content. CiTR's volunteers are tracking information, when it is available, from reel hubs and their boxes: tape speed, brand, date, recording length, notes on reel condition, accession number, content type, and any extra program and music details, in a spreadsheet. Unfortunately, record-keeping was not a priority for many of the radio hosts in the early years of the station's existence, and re-using tapes multiple times was common practice in radio. Often there is either no information at all associated with the tapes, or labels have overlapping information if the tape was recorded over. There is always a possibility that tapes could have been separated from their original box. Because of this administrative context, listening to the tapes is the only reliable way to assess their content (as well as whether or not they have Sticky Shed Syndrome). This is another significant, time-consuming challenge to the prospect of a digital reformatting project.

Copyright and intellectual property also present risks specific to radio archives. Musical performances are covered by a multitude of copyrights: mechanical rights, public performance rights, announcer rights, recording rights, lyrics/music rights. Public performance rights pertain to any on-air 'performance' including poetry readings, interviews, and lectures, among others. CiTR does not now, nor did it in the past, secure clearance agreements or authorizations from any of its content producers. Currently they pay SOCAN for broadcast performance and live stream rights, and CMRRA for mechanical royalties, which include keeping a single digital copy of albums for their digital library. Backup copies of digital libraries are not technically covered by this agreement. No agreements exist for podcasts at the moment, which many broadcasters make available for spoken word programming, but not music, because of legal liability.

According to CLIR's 2002 report, Building a National Strategy for Digital Preservation, "The laws and licenses that govern streaming were designed with the assumption that its use is ephemeral. It is unknown whether recording streamed audio for archival purposes is legal."[69] In a 2010 report, The State of Recorded Sound Preservation in the United States, CLIR warns "liability for copyright and common-law infringements may occur regardless of whether money

---

67   Lee, 195.
68   Library of Congress National Recording Preservation Plan, iv.
69   CLIR, 58.

is exchanged for sound recordings."[70] The Canadian copyright act specifies a term of fifty years from date of death of the copyright holder of a work, but provisions in the recently revised act allow libraries and others to make copies for backup, and in different formats. Is CiTR's desire to make historic recordings of their programming available to their listeners over the long term a possibility, given the legal climate and difficulty of identifying and tracing copyright holders to obtain permission? What about making their content available in jurisdictions outside of Canada, as web distribution allows? Will this project be for the purposes of preservation only? Will CiTR be able to secure funding for a full digitization project if long-term access is a question mark?

One thing is certain, the principle of benign neglect has worked in the past for paper materials, but digital records, if ignored, can rapidly become inaccessible. "Doing nothing reduces to zero, in a very short time, the possibility of preserving digital materials. One familiar example is failure to monitor and respond to the deterioration of digital media ... and the consequent inaccessibility of any data it might carry."[71] InterPARES research, which applies archival science and diplomatics to digital preservation, concludes that it is "literally not possible to preserve an electronic record like a record on paper." Therefore, digital preservation consists of preserving the ability to reproduce an electronic record in its proper form, with its essential digital components, rather than preserving the record itself.[72] We have to separate the technology from the digital records in order to render them using hardware/software combinations that are different from how those records were initially created.[73]

3.   Current Technical Practice for Digital Sound File Production at CiTR

CiTR is digitizing a small amount of their reel-to-reel audio tapes using one of the station's Revox PR99-MkIII machines, as a test case to determine the feasibility of digitizing their remaining tapes in house or outsourcing them to an external vendor. They use a half-track quarter-inch deck built in July 1990, which runs at the standard 7.5 and 15 inches per second speeds, and also continuously variable between approximately 5.1 and 22.5 inches per second. The volunteer doing the digitization work realigned it to Magnetic Reference Laboratories Test Tape 21J205, such that it is calibrated to +6 dB as per Studer/Revox specifications. Tape speed has been checked against a homemade test tape (segments of exactly 225 inches plus or minus 0.02 inches separated by leader) and found to be accurate within 0.04 percent. The tape deck is connected via XLR cables to a Focusrite Scarlett 2i2 AD/DA convertor at 96 kHz/ 24 bit stereo, running into REAPER (proprietary digital audio workstation software) on a dual-core Intel Xeon 2.0 Ghz. The raw files are WAVs; these undergo FFT and spectrograph analysis, applying noise-reduction and speed, EQ, and phase correction, and the resulting files as lossless FLAC — 96KHz / 24-bit for music, 44kHz / 16-bit for speech.

For their born-digital content, currently CiTR uses the SAM Broadcaster application to stream their programs and produce reports for the CRTC. They use a custom build of Burli podcaster software to create mp3 podcasts that are stored on a server in their office and backed up on a second server. Those podcasts, consisting of xml files with each program episode appearing as an item with a link to its associated mp3 file, are on their website, http://www.citr.ca/index.php/schedule/shows/. Podcasts contain metadata of the show title, date, host, description, genre, and website or twitter feed. CiTR is also building an mp3 library of albums ripped in iTunes and tagged with mp3 tagger using the ID3 v2.4 de facto wrapper standard, and currently have over 54,000 files taking up 294 gigabytes of space on another server. CiTR chose the mp3 format over a higher quality, lossless file format because of space considerations. However, they do have some albums in the lossless m4a format that they are not using for broadcast. Metadata captured by CiTR for their digital library includes

---

70 Ibid., 115.
71 Harvey, 124.
72 Duranti and Thibodeau, 20-21.
73 Harvey, 78.

basic information automatically generated by iTunes, with Cancon and femcon fields manually added. For more rare recordings not recognized by iTunes, all fields are manually added by the dj using mp3 tagger.

4.   Recommendations

While CiTR is not an archival institution, I believe it can employ some archival concepts and theory to positive effect with regards to the production of their digital audio file production. Considerations of accuracy, authenticity, authentication, and reliability are key concepts in digital preservation, and are defined in InterPARES' Creator Guidelines. For records to be authentic, identification of their means of production, manipulation, preservation, allowed access, identity, and integrity needs to be explicitly documented in a standardized and interoperable way. "Authenticity is, in fact, the result of a process, it cannot be evaluated by means of a boolean flag, and it is never limited to the document itself but extended to the information/document/record system."[74] This has an implication on file format selection, metadata schemas, and processing and storage decisions, among others.

Therefore, digital preservation ideally starts at the time of a record's creation. To create authentic copies of existing records, CiTR needs a project plan specifying policies and procedures, documenting the system environment as to exactly how the sound is captured, any changes made to files, file organization and metadata instructions, procedures to prevent, discover, and correct loss or corruption, and protection from media failure. This should be separate from a born-digital records plan. The recommendations mentioned below are introduced primarily with their large digitization project in mind, although preservation concepts apply to both. I hope they can be adopted and adapted with the same DIY spirit in which CiTR first approached radio programming. Given the fragility of magnetic media and the uniqueness of CiTR's collection, reformatting work will ideally be done only once with ongoing attention to preservation to ensure the work lasts for future generations.

Many audiovisual organizations, including IASA and Indiana University's Sound Directions, recommend the Broadcast Wave format as the recommended file format for long-term preservation. Broadcast Wave files are widely used, non-proprietary, platform-independent, and uncompressed. In essence, the difference between the Broadcast Wave over the regular Wave format is the ability to include metadata elements in the file's structure. CiTR can do this using BWF MetaEdit, an open-source tool produced by FADGI (http://sourceforge.net/projects/bwfmetaedit/). However, modifying a file by inserting additional metadata will adjust the checksum of the file, therefore a checksum hash value for the audio portion of the file only is recommended in addition to the hash value for the file as a whole for ensuring bit-level integrity of a file.[75] Checksums conducted on an ongoing basis using a command-line hash is standard procedure for ensuring the ongoing trustworthiness of a digital file as an aspect of its authenticity. Bressan and Canazza also recommend using JHove (Jstor Harvard Object Validation Environment), to automatically perform "format-specific identification, validation, and characterization of digital objects" (http://sourceforge.net/projects/jhove/).

While the provenance of their CiTR's audio records is not in question, researching the by whom, what, when, and how of poorly identified older programs is a daunting prospect. One whose importance must be underscored: "Recordings contain information on their artistic and cultural existence that goes beyond the audio signal itself. In this sense, a faithful and satisfying access to the audio document cannot be achieved without its associated contextual information, that is, to all the content-independent information represented by the container, the signs on the carrier, the accompanying material, and so on."[76] Once Digital Initiatives (UBC Library's digitization center) has finished digitizing *Discorder*, their in-house publication that

---

74   Bressan and Canazza, unpaginated.
75   FADGI, 25-26.
76   Bressan and Canazza, unpaginated.

has printed monthly programming schedules and articles since 1982, content research will be more feasible. Board member knowledge and the CiTR alumni Facebook group will be invaluable resources. CiTR's volunteers should be meticulous with the content-independent information, and it will be much easier to document information about the audio files being produced as they are made, rather than after the fact. CiTR has documented the system used for the test digitization work currently underway, and should endeavour to document its system as a whole, including custom alterations to software, as things change and work progresses. The InterPARES Guidelines for Creators will be a useful reference as work proceeds.

According to Harvey, the main types of metadata CiTR should consider including with their files are descriptive, structural, technical, administrative and preservation.[77] Duranti discusses integrity and identity metadata.[78] As a practical example for handling metadata, the archivist at the Western Front (a similar-sized non profit organization with comparable formats in their collection) is using a combination of PREMIS (which identifies preservation information) and Dublin Core to describe their audiovisual files. The main difference in metadata collected for the digitized versus the born-digital files is capture and processing information, which she obtains with the MediaInfo tool (http://mediaarea.net/en/MediaInfo). METS, or the Metadata Encoding and Transmission Standard, is another schema worth considering. FADGI is working on a metadata schema built expressly for audiovisual files, and that work should be monitored as it might become useful in the future, as might work being done on PBCore. Using standards for metadata like those described above improves the likelihood of interoperability with other schemas as they are developed, it aids future migration into new content management systems; metadata is, as Presto's *Techwatch* report clearly states, "absolutely crucial" for media asset management. A sample of recommended metadata elements for audio files from Bressan and Canazza is available here: http://www.hindawi.com/journals/jece/2013/489515/tab7/. Another example can be found in the user guide for the AVCC toolkit: http://www.avpreserve.com/wp-content/uploads/2013/10/AVCC_User_Guide_Beta.pdf. CiTR needs an appropriate, standardized metadata plan that works with their current CRTC-required data collection needs while also satisfying rights and preservation needs, keeping in mind that, "we must keep tweaking metadata systems through time — to calibrate them in accordance with our requirements and whatever happens to be the current technological environment."[79]

Audio files should be organized into logical groupings, with plans for file naming structures, categorization, classification, search and retrieval. AVPreserve released AVCC, a toolkit with guidelines, forms and a FileMaker database for cataloguing audio/visual material (which is only useful if CiTR already owns FileMaker Pro software). This is still in beta, but a useful example of how a large organization manages their audio content. One open-source option for a digital preservation content-management system is Archivematica. Though it is open source, it can require significant financial investment in set up fees for an individual organization's specific needs. As Archivematica works with its current clients, including MoMA and its audiovisual content, those custom builds will be released to the public, and may be worth keeping an eye on in the future. Once copyright details are sorted out, making content freely available on the Internet Archive is another possible solution that would cost nothing. Users are invited to freely submit audio content for streaming or download with a Creative Commons licence.

CiTR should keep multiple copies of their audio preservation files in different places, as well as access files that are used internally.[80] There is free, open-source backup and LTO management software from Bacula, used by Democracy Now, another non-profit media outlet. The Western Front uses LTO (Linear Tape Open) technology for their long-term storage, where they keep files and periodically check their integrity, and never overwrite them or re-save them. IASA and Presto Centre's *Techwatch Report* also recommends LTO for audiovisual

---

77   Harvey, 83.
78   Duranti, "Metadata and Archival Description".
79   Duranti, "Metadata and Archival Description.
80   IASA, 7.6.

material. However, this drive only has a fifteen to thirty year lifespan, so will need to be peri-odically revisited if selected as an appropriate storage medium, as technology changes. Cloud storage is another option that comes with its own pros and cons, one that AVPreserve has explored in detail for audiovisual materials. Once criteria is understood, e.g., the number of users needing to access storage, a budget and a plan for digitization, and we know approxi-mately how many reels CiTR would like to preserve for the long term and their accompany-ing space requirements, storage discussions can take place in more detail.

CiTR should formalize agreements with its djs and guests in the future, with appropriate legal advice, to obtain clearances to make broadcasted material available in different formats in perpetuity. For digitized content, volunteers should track down copyright holders as much as possible in order to obtain permission before making any content publicly available. Rights information should also be documented in the metadata of the digital files. Ideally, ongoing funding and staffing (like a recurring iSchool work-study position) could support the long-term digital preservation goals of CiTR. Building a database to facilitate classification, search and retrieval of digital audio records, for example, would be an ideal future Professional Experience position.

Organizations such as CBC Vancouver, the City of Vancouver Archives, and the Western Front have all been very supportive of a DIY digital audio preservation endeavour like the one CiTR is undertaking, providing advice, recommendations, and even equipment. With that support, and CiTR taking what it can from professional best practices and current research in the field from projects such as InterPARES, the station is well positioned to trail-blaze new and affordable routes toward long-term digital preservation of its historic and contem-porary audio programming.

5.   Bibliography

Bradley, Kevin, ed. "IASA TC-04 Second Edition. Guidelines on the Production and Preservation of Digital Audio Objects." South Africa: International Association of Sound and Audiovisual Archives-IASA Technical Committee 4, 2009. www.iasa-web.org/tc04/au-dio-preservation.

Bressan, Frederica and Sergio Canazza. "A Systematic Approach to the Preservation of Audio Documents: Methodology and Software Tools." *Journal of Electrical and Computer Engineering*. 2013. http://dx.doi.org/10.1155/2013/489515.

Brylawski, Samuel. "Preservation of Digitally Recorded Sound." *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*. Council on Library and Information Resources. 2002. http://www.clir.org/pubs/reports/pub106/pub106.pdf.

Casey, Mike and Bruce Gordon. "Sound Directions: Best Practices for Audio Preservation." 2007. http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/sd_bp_07.pdf. Accessed February 23, 2014.

Council on Library and Information Resources. "The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age." 2010. http://www.clir.org/pubs/reports/pub148/pub148.pdf.

Council on Library and Information Resources. "The Library of Congress National Recording Preservation Plan." 2012.

Duranti, Luciana. "From Digital Diplomatics To Digital Records Forensics," *Archivaria* 68: 39- 66. 2009.

Duranti, Luciana and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* 6(1): 13-68. 2006. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_02.pdf.

Duranti, Luciana. "Metadata and Archival Description." Powerpoint presentation, ARST 555, 2014.

Duranti, Luciana. "The Preservation of Digital Records." Powerpoint presentation, ARST 555, 2014.

Federal Agencies Audio Visual Digitization Working Group. "Assessment Report and Initial Recommendations." http://www.digitizationguidelines.gov/audio-visual/documents/AVPS_Audio_Metadata_Overview_090612.pdf.

Federal Agencies Digitization Guidelines Initiative (FADGI). "Embedding Metadata in Broadcast WAVE Files." 2013. http://www.digitizationguidelines.gov/guidelines/digitize-embedding.html.

Hackett, Y., William Underwood, and Philip Eppard. *Focus Task Force Report*. InterPARES 2. 2008. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_1_focus_task_force.pdf.

Harvey, Ross. *Preserving Digital Materials,* Second Edition. Berlin: De Gruyter Saur. 2012.

InterPARES 2 Project. *Creator Guidelines, Making and Maintaining Digital Materials: Guidelines for Individuals.* http://www.interpares.org/ip2/display_file.cfm?doc=ip2%28pub%29creator_guidelines_booklet.pdf.

Lee, Brent. "Issues Surrounding the Preservation of Digital Music Documents," *Archivaria* 50: 193-204. 2001.

McLellan, Evelyn, "InterPARES 2 Project - General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation." 2006. http://www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf.

Presto Centre, "TechWatch Report 01: AV Digitization and Digital Preservation." 2014. https://www.prestocentre.org/system/files/library/resource/techwatch_report_01_2014.pdf.

Weisz, Iolanda and Harry Hertscheg, CiTR. *50 Years of UBC Radio: 1937-1987*. 1987.