The C-SPAN Video Archives: A Case Study

Robert X. Browning

ABSTRACT

The C-SPAN Archives is a successful video archives of primary public affairs programming created and aired by C-SPAN. Now in its twenty-sixth year, the archives has confronted the challenges of creating an off-satellite archives, developing an indexing and cataloging system, providing access, and implementing new technology. These are issues with which other news archives must deal. This case study demonstrates lessons learned and problems solved that can be valuable for others thinking about these issues. As a total electronic collection of two hundred thousand hours, it represents a modern digital archives and a model for other news networks.

KEY WORDS

Audiovisual Archives, Archival History, Digital Preservation, Technology, Online Collections

The C-SPAN Archives, while unique in its development and accessibility, can provide some general lessons for others interested in preservation of and access to video. It also demonstrates how public affairs events can be made accessible through indexed, transcribed, and integrated audio-video digital records. It is a modern digital archives where citizens have easy access to the statements and actions of their leaders. It is also a model for what other news networks could do to open access to their collections.

Television is an important modern news and information medium. Yet it is ephemeral and thus poses unique challenges for both archivists who want to preserve television and to scholars who want to study it. The C-SPAN Archives has confronted these challenges and developed solutions that could be of value to archivists and scholars alike. In this case study, I present the challenges and solutions so that others might learn from its experiences.

There are two models for archiving television. One is the deposit model. Under this model, a local television station or national network deposits its older tapes with a library or museum. Often this happens when the station changes tape format and no longer wants to keep the older format. This can cause problems for the receiving institution because the technology is older or obsolete. These are usually production formats that can be expensive to maintain. The archival institution faces the same challenges as a paper archives—it must describe the records received. Issues of preservation acerbate the challenge—tape will not last forever, and the machines necessary to view the content need to be maintained. Many deposit television archives will make reference copies for patron use and preserve the originals in a temperature-controlled environment. However, though this process constitutes best practice, it is time consuming and expensive. The challenges of indexing are not much different from those of paper collections. There are always cases when a researcher will be looking for some information that is not indexed. For example, certain researchers are greatly interested in the buildings or backgrounds where interviews take place, but such context may not be considered important when the materials are first classified.

The second type of archival collection is the focus of this article. It is an off-air or off-satellite archives where an institution records the broadcast product of a television station or national network to preserve it.[1] We take it for granted that copies of newspapers are received and preserved in libraries, but the same is not true of the published products of television stations and networks.[2] After transmission, most broadcasts are lost to the public. They may be preserved by the stations or networks themselves, but historical researchers must wait and hope that the broadcaster will eventually donate these copies to a deposit archives as discussed above. Many times, stations archive only the individual stories and not the live news program on which they appear.

An off-air or off-satellite archives will record the daily newscast of a television station or national news broadcast and create a same-day record for the public to review. Off-air archives solve a particular problem, but create problems of their own. In the balance of this article, I will discuss the challenges, solutions, and possibilities of off-air archives and the C-SPAN Archives, in particular, one successful off-air archives.

C-SPAN is a nonprofit, national television network based in Washington, D.C., that televises the U.S. House and Senate, presidential events, congressional committees, news conferences, political conventions, political speeches, and other political events. C-SPAN coverage is unusual because it broadcasts entire events without editing or commentary. As such, it is a primary source record of the debates, deliberations, and pronouncements of U.S. public policy. C-SPAN coverage creates audio and video records of government in action. Nonetheless, it is ephemeral television that can quickly disappear.

In 1986, C-SPAN's founder and chairman Brian Lamb met with a small group of faculty at Purdue University, his alma mater, to discuss how C-SPAN programs could be used in teaching and research. The faculty indicated that this was unlikely without an archives to organize and preserve the materials. C-SPAN had a limited production archives at the time, but erased most tapes to reuse them. Lamb indicated that the small network had limited resources to create a research archives and suggested that the faculty at Purdue do so.

One visible model existed at the time: the Vanderbilt (University) Television Archives. Established in 1968 in the Vanderbilt University Library, the archives recorded the half-hour nightly evening news programs, indexed them, and loaned copies to researchers. The Library of Congress also recorded evening news programs, but those copies were not readily available for use. Vanderbilt operated under a clause of the 1976 U.S. Copyright Act that permitted libraries to record news programs and to loan copies as long as they did not sell them. The enactment of this clause settled the 1973 lawsuit, *CBS vs. Vanderbilt*, in which CBS asserted that it owned the copyright to its broadcasts and that Vanderbilt University was in violation of that copyright by operating the television archives.[3] The law allowed the lawsuit to be settled, and Vanderbilt continued to operate. Essentially, the library could save television news programs in the ordinary course of its operation and loan copies much as it would loan other materials. Later, the Vanderbilt News Archives faced financial problems. Financial reorganization, a new fee structure, and subsequent grants ensured its future. The television archives is now forty-five years old.[4]

Following this model, in 1987, Purdue University created the Public Affairs Video Archives to record, index, and distribute all programming telecast on the two C-SPAN networks, C-SPAN and C-SPAN2. The difference between Vanderbilt and Purdue is that C-SPAN granted Purdue a license to make duplicate copies.

This represented the first licensed television archives in the country. Purdue did not need a license to record and index the programming, but the license permitted the Purdue archives to distribute copies of videotapes to educators for a fee. These fees would be used to support the archives.

The technology used and the story of how the archives came to win a 2010 Peabody Award for its efforts are interesting but not the focus of this case study.[5] Rather, I concentrate on the archival issues involved and how the experience of the C-SPAN Archives can be used to understand the requirements for archiving and preserving modern electronic news broadcasts.

I discuss four key archival issues. First is the challenge that off-air archives present for universities, libraries, and museums. Second are the issues of indexing and cataloging that are so important for any collection. Third is the importance of providing access. Fourth is the challenge of technology that off-air archiving must confront.

It is natural for universities, libraries, and museums to be interested in preserving local and national television because it has parallels to the preservation of newspaper collections.[6] Local television creates a valuable resource that fosters the study of local history. However, recording off-air television differs from what local history archives normally do. Universities and colleges are more appropriate venues for these endeavors than local historical societies because they have audiovisual departments and their libraries traditionally handle videotape viewing for students. The equipment needed for an off-air archives would be newer and thus simpler to maintain than older production equipment that comes with deposit collections. The collection in an off-air archives is the air product of the station or network, not the individual stories, the variations of the stories for different broadcasts, or the related "b-roll" video that did not get to air.[7] The difference would be similar to the files of the newspaper reporter that include the notes or tapes from the original interviews as well as the published newspaper story.

These archives face the issues of financing and supporting themselves. Unlike deposit archives, which can wait until funds are available to process a new collection, an off-air archives faces immediate ongoing expenses for tapes (or increasingly for digital equipment) as well as ongoing staff support to process the incoming video. Repairing and keeping up with changing technology also incurs costs. Again, a university that can call upon its audiovisual department for support has options to defray some of these costs.

The second challenge is indexing and cataloging. While archives are strong in the area of appraisal, archival description, preservation, and reference, there are no agreed-upon standards for indexing and cataloging electronic news coverage. Classification by date, network, and location is straightforward, but collection-level indexing does not get to the item level. Leaving it to the patron

to scan hours of video is not always a desirable course, and most institutions would like to provide more assistance in getting to the day of the event. For national news, indexes to national newspapers can lead users to the correct day, thus simplifying the research process. For local news, it depends on whether any index to the newspaper exists to allow a researcher to pinpoint when an event took place and thereby access the day upon which it occurred.

Program-level cataloging requires the development of a set of keywords, names of individuals and organizations featured, and topics reported.[8] While some groups have worked on this problem, no uniform set of standards or software exists.[9] It is an interesting problem for librarians and archivists. At the same time, developing one's own standards creates uncertainty about whether the correct approach is applied and reluctance to become a pioneer in this area.

Two goals, preservation and access, guide these archival projects. They are not always complementary, in that one way to preserve the records is to limit access. Using an original recording in the analog format risks damaging the medium. Thus, reference copies are made. This is less true in the digital era when duplicate copies can easily be made in less than real time. In fact, the digital era can also allow for the expansion of access.

Technology offers the greatest challenges and opportunities. It was relatively easy to use videotape and video recorders to capture off-air content. The technology was known. The tapes would pile up, but they could be labeled, stored, and played back. The path to digital recording, while it offers great opportunities for storage, access, and preservation, contains many variables to consider. Many digital video recorders exist for short-term capture and playback and do not have options for getting materials off for long-term storage. Some recorders have CD or DVD capability, but the standards are elusive and this technology should be approached carefully when the longer term is under consideration. Computer recording capability using video capture cards and CDs offers more flexibility and will be explored below.

With the four issues identified, I now turn to how C-SPAN addressed them beginning in 1987 and what can be learned from that experience. The history of the C-SPAN Archives can be divided into four periods: 1) the pre-archival period, 1979 to 1987; 2) the university analog period, 1987 to 1998; 3) the digital corporate period, 1998 to 2010; and 4) the post-2010 Video Library period. Each has its distinct characteristics and development.

Pre-Archival Period, 1979 to 1987

C-SPAN began televising on March 19, 1979, when the U.S. House of Representatives initiated a video signal of its proceedings. It was a small operation with four employees and a graphics machine to periodically insert

network-identifying "crawls" across the bottom of the television screen. There was no other programming or equipment to record what was transmitted. Even the House recycled its videotapes so that no tapes exist of proceedings from 1979 to 1982, when the House started depositing tapes with the Library of Congress.[10]

In 1980, C-SPAN began airing call-in programs and National Press Club Speeches, and, in 1981, it covered the confirmation hearings for President Reagan's cabinet on Capitol Hill. Most of these tapes were reused to record new material so that few recordings remain available for this period (see Table 1). These were originally recorded in 3/4-inch production tape format.

Table 1. Digitized Tapes Extant from the Early 1980s

1979	9
1980	56
1981	54
1982	210
1983	698
1984	1,619
1985	1,706
1986	3,210
1987	2,296

These tapes sat for years in the C-SPAN Tape Library in Washington, D.C. In 1988, they were given to the National Archives in a deposit agreement. The idea was that they would be kept close to C-SPAN's Washington headquarters for retrieval and would also be accessible to the public. Initially, the National Archives returned the call-ins that it did not think had much archival value, but retained the hearings and speeches. The National Archives sought to terminate the agreement and return the balance of the tapes to C-SPAN in 2007. The C-SPAN Archives took possession of these materials and indexed and digitized them. This demonstrates the problem of managing obsolete formats in a deposit archives.

The University Analog Period, 1987 to 1998

The second period began in September 1987, when the Purdue University School of Liberal Arts established the Public Affairs Video Archives as a licensed, off-air university archives to record, index, archive, and distribute to educational institutions all programming telecasts on the two C-SPAN networks.[11] The founding motivation was that C-SPAN events in their entirety, including unedited content, were valuable records of public affairs and should be captured and

preserved. C-SPAN donated two satellite dishes and granted the license to distribute video copies for a fee, but the university was completely responsible for organization, operation, and funding. The license permitted the sale of duplicate copies to educational institutions, and the fees would defray the costs of operating the archives.

Purdue, under the leadership of David A. Caputo, then dean of the School of Liberal Arts, saw the off-air, university-based archives as a way to make these materials available to current and future scholars. Over the years, few scholars actually visited the archives, but almost five hundred thousand tapes were distributed to professors, teachers, researchers, and libraries to be used in teaching and research at all levels.

This period marked a time of development for the archives. Cataloging standards were established, as well as the technology base that allowed the archives to grow and flourish. Computing was becoming decentralized, so it was easy to buy personal computers and database software to develop the initial database of C-SPAN programming. One of the earliest employees was a first-year computer science student at Purdue who wrote the original programs and designed the initial database. Those database fields are still being used twenty-five years later. They include the category of the program, the format of the event, the program name, an abstract, the sponsoring organization, and the location. A librarian brought into the project during the first year developed a nested set of keywords based on Library of Congress subject headings. Since C-SPAN programming is all public affairs event coverage, the scope of the content was manageable, albeit large. Figure 1 shows the arrangement of a physical record created for a single event.

Program Category	White House
Program Format	News Conference
Program Title	Presidential News Conference
Program Date	January 13, 2013
P**rogram Sponsor**	White House, East Room
Program Location	Washington, DC
Program Abstract	President Obama held a news conference in the East Room of the White House. Among the topics he addressed were efforts to implement gun control measures following the school shooting tragedy in Newtown, Connecticut, and the need to raise the debt ceiling in order for America to "pay its bills." In his remarks he called congressional Republican refusals to raise the debt ceiling "irresponsible" and "absurd," saying "they will not collect a ransom in exchange for not crashing the American economy."
Program Person	Obama, Barack, President, United States, 2009–
Program Tags	Business & Commerce -> Budget, Federal -> Federal Budget Courts & Judicial Process -> Crime -> Gun Control

FIGURE 1. This detail shows the format of a C-SPAN archival record.

Foreign Affairs and Defense	**Foreign Policy**	Communism
		Democracy
		Developing Countries
		Environmental Policy
		Foreign Aid
		Humanitarian Issues
		Immigration
		International Summits
		Iran-Contra Investigation
		Peace Corps
		War Powers Act
	International Security	[More keywords]
Health and Welfare	Drug Policy	[More keywords]
	Health Care Reform	
	Health Policy	
	Social Policy	

FIGURE 2. This sample illustrates a portion of the hierarchy of the archives' keyword system.

All the programming at the archives originates from one network as recordings of public affairs events, so noting physical formats, genre, producers, and rights was largely unnecessary. Instead, archivists could concentrate on the content. A sample of the keyword system is found in Figure 2, and a full listing of these tags can be found on the website.[12] Table 2 shows how these database fields and keywords are incorporated into the workflow processing of video.

The audio-video department at the university provided the technical support to record the C-SPAN programming. This was a great help because it allowed C-SPAN Archives staff to work on the indexing and fulfillment functions without having to worry about the technical operations. Over time, when the staff grew, they took over the technical aspects and integrated them with the software to make a sophisticated machine-controlled recording system.

During this university phase, financing came primarily from fees charged for duplicate tape copies and initial subsidies from the college and from C-SPAN. The need to generate revenue led to some marketing efforts. The archives staff developed tapes that might be popular for teaching certain subjects or that were compilations of short events for classroom use. Flyers highlighting these tapes were prepared for various academic disciplines and mailed to lists acquired from professional associations. The database was used to develop these flyers automatically in a print format. Catalogs of the entire holdings were published on quarterly and annual bases and sold to schools and libraries. Demand for programs that had just aired was constant, but these sales were usually in ones and twos. Some felt that promoting the use of the archives through tape sales was too commercial for an archives, but there was ongoing pressure to realize

Table 2. Workflow of Indexing and Storage

Step	Process	Explanation
1	Skeletal records received electronically from C-SPAN production system.	These records have a program name, an internal ID number, a network, and an airing time.
2	Archival program fields outlined in Figure 1 are entered.	Category, format, title, sponsor, location, abstract, people, and keywords
3	Pictures attached.	JPEG images of each person as well as opening, closing, wide, and audience shots are stored with each program record.
4	Times set.	When the program has aired, the exact time that the program began and ended are entered. These times are entered by clicking on a time-stamped picture that enters the time into the database.
5	Master checked.	Air records are checked visually, to make sure that the longest airing was selected as the master record.
6	Digital copy moved to archival tape robot.	A Sun StorageTek digital tape robot is used to store all first-generation video.
7	Digital copies are transcoded into multiple formats that can be played on the Web.	Linux multiple processor servers transcode the different formats using FFMPEG software.
8	A Web format copy is transferred to the cloud.	All video is stored at Amazon Web Services cloud architecture for redundancy. Highly demanded video can be played from this platform.
9	Second copy of first-generation video is transferred off site.	Every several months, copies of the primary tapes are transferred to remote underground storage.

enough revenue to cover costs. C-SPAN assisted by providing tape grants to professors who could then use these coupons to purchase materials. Purdue would be paid from the redeemed coupons.

Over time, C-SPAN and the university realized that the archives had grown in size and complexity and a new governing agreement was needed. This agreement acknowledged the complexity of the intellectual property ownership as well as the archives' need for revenue. It brought a new level of financial stability by allowing the archives to sell to all viewers, whether educational or not, and provided for C-SPAN to collect a share of that revenue after the archives had achieved a certain base revenue needed for operations. It also allowed C-SPAN

to purchase equipment needed for duplication that the archives would use, but that would be titled to C-SPAN. This allowed the archives to grow sufficiently to handle increased tape duplication and phone calls inquiring about materials. In essence, C-SPAN recognized that the independent, university-based archives could provide a valuable resource to the network and its viewers through the answering of questions about what it aired, by providing copies of aired programs to viewers, and by creating an accurate log of the network's programming. This agreement in many ways provided a model for the relationship between a university off-air archives and an independent network.

The Corporate Archival Period, 1998 to 2010

The third phase in the history of the C-SPAN Archives came in 1998 when the archives separated from Purdue University and became a part of C-SPAN. A change in university administrators resulted in a reluctance to support a university-based archives that depended on sales to support itself. Therefore, C-SPAN felt the only choice was to take over the entire operation. The archives was successful in terms of recording, indexing, preserving, and distributing content, and C-SPAN recognized this value to the company. The university expressed interest in the collection and its research potential, but had no other means of supporting it besides depending on sales revenue. People would always want copies of materials, and charging for those copies was a viable way to support the archives.

So on July 1, 1998, the C-SPAN Archives organizationally and physically separated from Purdue University and became an operating unit of the C-SPAN television network. C-SPAN also had a production tape library in its Washington, D.C., headquarters that maintained tapes used on-air. Thus, C-SPAN was the first network to support both a production archives and an off-air research archives. By doing so, it became the first network to preserve every hour of every broadcast day.[13]

Moving the operating archives was no easy task as it replicated all the services it had performed at the university while continuing to record two networks 24/7, indexing all the programming, and serving the viewers who continued to want copies of the programs. Some in the university were disappointed that the archives was no longer located on campus, and others felt that the university lost direct control of a valuable resource.

C-SPAN spent money to purchase the intellectual property that the university owned and to invest in equipment needed to keep the archives modern and efficient. This third phase is marked by C-SPAN's investment in technology that created the award-winning archives so publicly accessible today. Initially,

duplicating videotapes was replaced by duplicating DVDs. DVDs became the distribution medium of choice for individuals, and thirty thousand DVDs replaced the same number of VHS tapes being duplicated and shipped from the archives.

Some staff moved from the Purdue University payroll and became employees of C-SPAN. Certain things not possible in the university environment became possible. These included allowing orders to be placed with credit cards. While the entire structure in place at the university had to be replicated, things could be done more quickly in the private environment.

During this time, the transition from analog to digital was realized. C-SPAN was willing to invest in the digital equipment necessary to convert recordings from videotape to digital storage as well as to provide the additional equipment to digitize 120,000 hours of analog tapes. In 2002, the archives purchased two Grass Valley servers along with a StorageTek digital tape robot to begin the digital recording.[14] Since most broadcast-quality equipment is designed for commercial television production, it took some work to use it for archival recording. Since the beginning, the archives had software development capability, so they were able to write software to enable the digital recorders to mimic the analog system. Every two hours, a new file was started, and, similar to the analog tape system, the files were named by the exact time that the file started. With the new equipment, however, they were recording the time down to the second. A second machine recorded a backup copy simultaneously. One machine was configured with six inputs and two outputs, and the second machine, primarily used for duplication, was configured with two inputs and six outputs.

Recording was 10 Mbps in long GOP MPEG2 format.[15] Because of the size of these files, a separate proxy recording was designed. Using PCs initially and later using specialized video recorders, the archives recorded a second MPEG1 video stream with the timecodes synced so it could be viewed at the desktop for cataloging purposes. Finally, a third recording stream was created using another set of PCs to record the video in Real Video format for streaming on the Web.

To assist the catalogers with scanning large amounts of video, still images were captured every six seconds and displayed on contact sheets. Each of the images could be clicked on to open it full screen, and each image was time-stamped to match the moving video in the different formats. A software tool was developed to allow the cataloger to set the beginning and ending times of programs by clicking a button below the picture. This prevented transposition errors when entering times down to the frame of video. Figure 3 provides a sample of an issue.

Using this tool, the catalogers are able to select and archive five pictures with the program record. These are the opening and closing shots, a wide shot, an audience shot, and a marketing shot to be used whenever the program is displayed. Then, as they catalog the program, they save a shot of each person.

FIGURE 3. A saved image from a White House program demonstrates the labeling of a still image.

These pictures allow the catalogers to quickly see the essence of every program without accessing the moving image. In addition, the pictures of individuals allow the catalogers to avoid linking the wrong person to a program when they need to determine which of two persons with the same name is appearing. People often appear in one capacity and are then elected to another office or represent a different group. The picture allows another level of information in the cataloging process. It also documents how people change over their public lives. Over seven hundred thousand pictures have been indexed in the database.

Also during this period, digitization of the analog collection was completed. There were 87,000 VHS and SVHS tapes that held 120,000 hours of first-run analog content. Manually ingesting all of these tapes would take tens of thousands of hours, so the archives looked for ways to automate the process. They purchased a VHS robot that could hold eight VTRs and a hundred tapes at a time. Since all the tapes were barcoded and time indexed, the software could read the barcode and create a corresponding new file record in the database. Each tape and file record was linked to the content record that contained the information about what was on the tape.

C-SPAN telecasts roughly eight thousand hours of original content per year across three networks twenty-four hours a day, seven days a week. The recording system would use a new tape or create a new digital file every two hours and label it with the date, time, and network. Tapes that did not have original content were not retained in the initial years. In the ingest process, the tapes were scanned and checked against the database to ensure that they contained master content.

Before the robot was installed, the archives pulled significant videos from the collection and digitized them manually. These included Brian Lamb's *Booknotes* interviews and the Robert Bork nomination. Once the robot was in

FIGURE 4. This tape robot was used to digitize archival content.

place, staff began with the newest content, reasoning that the oldest tapes, which were not quite twenty years old, were good quality and were stored in climate-controlled conditions and therefore not under threat of degradation. The staff knew that the quality of the indexing was better for the later years than the earliest years. So, they concluded that the digitization would go faster if they proceeded with the newest materials first, which contained less duplicate content and partially cataloged materials.

Each day the robot was loaded with a hundred tapes, and all day and night it played the content into a new server purchased just for digitizing. It took almost twenty-four hours to play all the tapes back into the system. A second robot was added when the digital recording systems were switched to a new format. The digitization was equivalent to sixteen additional networks being recorded twenty-four hours a day. The archives hired a student to load the robot on the weekend and was soon digitizing two hundred tapes a day, twenty-four hours a day, seven days a week. One hundred twenty thousand hours were digitized in about three years.

During this time, the archives moved to its second generation of digital recording equipment, switching from Grass Valley 10 Mbps MPEG2 format to 6 Mbps MPEG4 format. This equipment produced much smaller, more flexible, higher quality files. Archivists were able to transcode the files in Web-viewable quality with off-the-shelf Linux servers and could quickly make the files available to the public for viewing. Four encoders were dedicated for the three C-SPAN networks and sixteen encoders for the analog materials. For redundancy, two servers created lists for recording the materials simultaneously on two different servers. Later, three HD encoders were added, and they encode the three networks in HD format at 9 Mbps.

Because of high interest in programs immediately after they air, the archives added a record list to capture five-minute files. This meant that the recorded content could be played about ten minutes after it started airing, instead of two hours later. These files were linked so that they played continuously. After the one-hour files were available, they replaced the five-minute files, and the five-minute files were erased. The one-hour files were archived continuously to the digital tape robot and transcoded into Web-viewable files.

The original recordings are archived in 9 Mpbs HD quality. This format is used to make DVDs and for licensing. The Web-viewable H.264 formats[16] vary from 50 Kbps to 2 Mbps. Because the C-SPAN Archives is primarily a research and educational archives viewable on the Web, these formats are sufficient. These formats are also pushed to the Amazon Web Services cloud[17] for redundancy and to enable playback of high-demand files from the cloud.

The period 1998 to 2010 was a time of growth as C-SPAN invested in technology and the technical staff realized a large number of firsts. The system

was stabilized with few technical problems, and the archives moved into larger facilities with very good power and power backup. During this time, staff continued to refine the cataloging, ported the database to MySQL,[18] and continued to process video on a daily basis without a backlog.

The Video Library Period, 2010 to the present

The C-SPAN Archives is currently in its public Video Library phase, which began in 2010. During this phase, all the recording, indexing, and technology came together to create the C-SPAN Video Library.[19] Officially launched on C-SPAN's anniversary, March 19, 2010, the Video Library allows users to search, watch, clip, and share any program in the now more than 200,000-hour collection. The Video Library received many accolades on release, including the assertion that "it will redefine social studies education."[20] In 2010, the C-SPAN Video Library received a Peabody Award, the highest award in journalism.[21]

Refinements continue during this time, including the creation of *Congressional Chronicle*.[22] Speaker indexing was added to the congressional proceedings, and members' words are linked to the printed *Congressional Record*. This was a first and is the only source that records what members actually say on the floor. Members are permitted to insert their own remarks into the *Congressional Record* and do not have to be present and actually speak. Now the *Congressional Chronicle* exists as a record of what actually happened in the House and Senate. Using the *Chronicle*, researchers can search and play any speaker's remarks for any day that the Congress was in session.

Since 1994, the archives has been capturing the closed-captioning text on seven networks, including the three C-SPAN networks. Technology staff developed indexes and linked every word to the recorded video. The release of the C-SPAN Video Library enabled text searching to access the video point where the words were spoken. This feature combined with the archives' own weighted search of all the text and indexes created a very powerful search engine and indexed video collection.

Another refinement during this period has been the development of within-program indexing. Speakers in the electronic records C-SPAN created during the recording process can be matched to the archives' video record. The result was a speaker-by-speaker index within the program for a large number of new programs in the Video Library. Now, a researcher can move right to a particular speaker during a four-hour hearing. The captioned words are also matched by time, so one can also read and search the text. This was a powerful addition to the archives' item-level indexing.

For presidential events, the White House transcripts were linked to the videos. This creates the same level of word linkage, but uses the actual transcripts

rather than the captioned text. The index entry (audio, video, and transcript) created for presidential events provides an important historical record and the first of its kind.

The C-SPAN Video Library is widely used. On average it receives about a hundred thousand video views per day along with around twenty thousand unique programs views. Programs are embedded in sixty-five thousand websites a year. The C-SPAN Archives has truly moved from the idea of preserving and making C-SPAN content accessible to a very accessible, widely used video collection.

Lessons Learned

What lessons can others learn from this case study? First, off-air archives are a valuable resource for preservation of and access to broadcast news and public affairs recordings. No other easily accessible record of these broadcasts exists. News specials and documentaries will always be kept, but the daily, routine local news broadcasts and complete records of networks such as C-SPAN, CNN, MSNBC, and Fox News require off-air archives.

Second, a university, national archives, or library are the only organizations likely to have the resources and technical infrastructure to establish and maintain such an archives. A committed local library or historical society could attempt to archive local evening news broadcasts with today's technology and provide a valuable service.

Third, long-term institutional support is not always possible. The financial difficulties that Vanderbilt experienced over the years bear this out. A deposit archives is easier to maintain because it is not under the daily pressures of acquiring and cataloging new content; a deposit archives can wait until funds are available. The experience of Purdue and C-SPAN shows that administrators with different agendas, cost or space considerations, or different visions for what their institutions should support can threaten an archives' existence. Had the Purdue University Public Affairs Video Archives not created a significant collection, archival and retrieval systems, and a valuable database, C-SPAN might have gone the direction of other networks and relied on the transcoding of its production archives for preservation and access.[23]

Fourth, indexing and cataloging are essential. Few standards guide the indexing of off-air public affairs. Vanderbilt University and the C-SPAN Archives developed their own standards using authority tables and keyword vocabularies. Institutionalized over the years, they create a de facto standard. For a local collection, event record indexing by date and time provides a guide to the collection that can be matched to other indexed local news collections.

Fifth, technological investment and improvement are critical to the success of a video archives. This was begun at Purdue, a major engineering and scientific institution, where the technological base was established. Computerization began right away with the development and constant evolution of the database. The database did not just contain the index to the video records, but also an index to the physical location and tools to manage the retrieval and duplication. This evolved into tools for digital recording, duplication, and storage. Investments in technology are expensive. Initially Purdue and later C-SPAN were willing to make those investments at a critical time and made the modern, accessible archives possible.

Sixth, it is relatively easy to make an archives accessible on the Web. It is now easier than ever to create a public website to advise people of the existence of a collection. Since these websites are indexed by search engines, more people will become aware of the archival efforts behind them. The index, dates of the collection, pictures, and even the video itself can be posted. Computer software is easy to access today to build summaries of the collection or even a database itself. These records can be exported to the website with ease. Of course, a video's copyright is an important issue to negotiate. Some interpret the clause under which Vanderbilt operates to permit online access, but it is uncertain if everyone agrees on that point. No one wants more litigation, so for a local archives, negotiating the terms of access, including online access, is important.

These lessons are important and demonstrate the value as well as the challenges of creating a video archives. They apply to the technologies coming into existence every day. Archivists are aware of the value of web pages and Twitter feeds and are taking steps to archive them on an ongoing basis so that information will not be lost as it was in the early days of television. Modern search engines make external indexing less necessary with these data streams. As technology advances along with an awareness of history, contemporary archival practices will need to be designed to preserve the information being generated. At the same time, the new challenge will be preserving the sheer volume of the information created.

Notes

[1] Cable television is not transmitted through the air, but via satellite. In this article, I treat off-air and off-satellite recording the same and refer to them as "off-air recording" for simplicity.

[2] Local stations are broadcast entities that transmit signals through the air. National networks can be either broadcast or satellite.

[3] See Paul M. Gherman, "The Vanderbilt Television Archive," *Perspectives on History* (October 2004). A very thorough account of the origins of the Vanderbilt Television Archive and its legal struggles with CBS can be found in Lucas Hilderbrand, *Inherent Vice: Bootleg Histories of Videotape and Copyright* (Durham, N.C.: Duke University Press, 2009).

[4] Marshall Breeding, "Digitizing Brings New Life to Video Collections," *Computers in Libraries* (October 2008).

[5] For a discussion of technical issues, see Robert Browning, "Creating an Online Television Archive, 1987–2013," *International Journal of Digital Curation* 9, no. 1 (2014).

[6] Debora Cheney, Jeffrey Knapp, Robert Alan, and Pamela Czapla, "Convergence in the Library's News Room: Enhancing News Collections and Services in Academic Libraries," *College and Research Libraries* (September 2006).

[7] One definition is "B-Roll is the footage that doesn't involve either an on camera talent or an interview face being on the screen." Kris Simmons, "What Is B-ROLL Footage?," *6strong*, http://www.6strongmedia.com/what-is-b-roll-footage.

[8] Program-level, or item-level, indexing gets one to the original video as broadcast. One must still scan the entire video to find the portion that one is interested in. CSPAN's goal is to go beyond the program-level and index within the program so that one can find words that were said in the video.

[9] See Mike Cox, Linda Tadic, and Ellen Mulder, *Descriptive Metadata for Television* (Burlington, Mass.: Elsevier, 2006).

[10] A 1989 National Archives report indicated that "Videotapes of House Proceedings from 1979–82 are not extant." *Guide to the Records of the United States House of Representatives at the National Archives* (House of Representatives, 100th Congress, 2nd Session, Doc. No. 100-245, January 1989), 8.

[11] C-SPAN3 began in 1997 and the archives began recording it in 2001.

[12] C-SPAN, www.c-spanvideo.org/browse?browse=tag.

[13] In his history of film and television archives, Sam Kula observed, "No network retains every hour that is broadcast in the form that it was originally transmitted, complete with advertisements." Sam Kula, *Appraising Moving Images* (Lanham, Md: Scarecrow Press, 2002), 19. This is not true; C-SPAN does.

[14] For information about Grass Valley video servers, see www.grassvalley.com. StorageTek digital storage robots are now part of the Oracle Corporation, "Oracle Tape Storage," http://www.oracle.com/us/products/servers-storage/storage/tape-storage/overview/index.html.

[15] This Library of Congress site contains details on the various formats for digital video and digital preservation, "Sustainability of Digital Formats Planning for Library of Congress Collections," http://www.digitalpreservation.gov/formats/fdd/browse_list.shtml.

[16] See Library of Congress, "Sustainability of Digital Formats."

[17] Amazon Web Services is a large provider of remote computing services, known as "cloud computing." An introduction can be found on its website at http://aws.amazon.com/.

[18] MySQL is a popular open source database that is very inexpensive to use, http://www.mysql.com/.

[19] C-SPAN, www.c-spanvideo.org.

[20] Frances Martel, "C-SPAN Online Archives Will Redefine Social Studies Education in America," March 16, 2010, www.Mediaite.com.

[21] Kirsten Gibson, "C-SPAN Archives Gifted with Prestigious Award," *The Exponent Online*, May 27, 2011, http://www.purdueexponent.org/city_state/article_c199c23e-87f8-11e0-833f-001a4bcf6878.html. See also, "Peabody Award Winners 2010: NPR, PBS, CNN, C-SPAN, 'Good Wife,' HBO and More Win Prestigious Awards," *The Huffington Post*, March 31, 2011.

[22] The *Congressional Chronicle* is contained in the C-SPAN.org website and contains information about members, floor actions, committee hearings, and press conferences, CSPAN, *Congressional Chronicle,* www.c-span.org/congress.

[23] C-SPAN continued to support the archives even as physical DVD sales declined. It permitted the C-SPAN Video Library to provide free viewing as part of its public affairs mission.

ABOUT THE AUTHOR

Robert X. Browning is an associate professor of political science and communication in the Brian Lamb School of Communication at Purdue University. In 1987, he became the founding director of the C-SPAN Archives. He is the author of *Politics and Social Welfare Policy in the United States* and papers on redistricting. Awarded the George Foster Peabody Award for its online Video Library in 2010, the C-SPAN Archives is housed in the Purdue Research Park where Browning directs the operations including the database design, indexing standards, and software development.