

Preservation and archiving of digital media

Clifford B. Anderson*

Associate University Librarian for Research and Digital Strategy, Vanderbilt University, Nashville, TN, USA

Abstract. This paper provides a brief history of the Vanderbilt Television News Archive that was established in 1968 with the goal of recording and preserving national news programming on the three major networks at the time (ABC, NBC, and CBS). The archive has faced several challenges as it evolved (it now covers representative news from the Fox and CNN cable networks) - most notably financial and legal issues – who really “owns” the news? Even today archiving digital news remains financially and legally challenged as the number of news networks increase and privacy laws emerge, resulting in the creation of “piracy archives”. The author also touches on the many ongoing issues that need to be addressed; e.g. the number of copies to be created, the importance of metadata, the technical requirements, what qualifies for preservation, and the ever-present issue of sustainability.

Keywords: Digital archiving, digital preservation, television news archive, piracy archives, preservation sustainability

1. Introduction

The preservation of digital media is a broad topic and, as an archival practitioner rather than theorist, my goal in this paper is to share my perspective on digital preservation as administrator of a long-running digital preservation project, the Vanderbilt Television News Archive. While my focus is on the preservation of television news, I trust that our experiences at the Vanderbilt Television News Archive provide a helpful case study of the challenges of digital media of all kinds.

The Archive started on August 5, 1968 as an experiment to record and to preserve national news programming on the three major networks. The story goes that Paul Simpson, a Nashville insurance executive, instigated the creation of the Television News Archive after seeing an interview on the evening news in 1967 with Timothy Leary. After discovering that the broadcast network could not proffer a copy of the tape, Simpson partnered with Frank Grisham, Director of the Joint University Libraries at Vanderbilt, to begin preserving and providing access to the past news broadcasts [1]. Robert A. McGaw, Secretary of the University, provided crucial institutional backing. During the same era, other institutions like the University of California Los Angeles (UCLA) and the University of Georgia began developing television archives as well [2].

Simpson represented the kind of sophisticated viewer who instinctively rejected what media theorist Herbert Gans subsequently termed the “mirror theory”, namely, the thesis that “events determine story selection, with journalists simply holding a mirror to them and reflecting their image to the audience” [3]. Simpson suspected that the perspective of New York-based news directors biased the national discourse

*Tel.: +1 (615) 322 6938; E-mail: clifford.anderson@vanderbilt.edu.

about the contentious cultural and political events of the sixties. Members of Congress shared the concern that news directors might be “rigging” television news programs for political reasons; Simpson testified at congressional hearings in 1972 on the possibility that news networks were “staging” news events [4]. But moving from suspicion to evidence required an archive.

Despite the odds, the Archive has persisted for more than fifty years, continuing to fulfill its core mission of preserving and providing access to the history of television news. The collection now includes fifty-two thousand five hundred hours of video content. Researchers come to our offices in Nashville from around the world to conduct research and our website provides offsite users with the ability to search and loan clips. If Walter Cronkite’s famous sign-off line was “And that’s the way it is”, the informal motto of the Vanderbilt Television News Archive is “And that’s the way it was”.

The preservation challenges of the Vanderbilt Television News Archive are daunting. The Archive maintains somewhere in the neighborhood of a petabyte of data across different formats as well as 1.2 million metadata records. The staff of the Archive works in concert with technologists in the library and at the university to preserve this information and also partners with the Library of Congress to keep a copy of the archive offsite.

2. Sustainability

Trevor Owens remarks, “If you want to evaluate how serious an organization is about digital preservation, don’t start by looking at their code, their storage architecture, or talking to their developers. Start by talking to their finance people” [5]. In like fashion, Edward Corrado and Heather Moulaison Sandy contend that “digital preservation is in many ways primarily a management issue” [6]. The question of finance has loomed large at the Vanderbilt Television News Archive during its fifty years of existence. From a bootstrapped operation, the Archive emerged into a large scale enterprise of thirteen staff members by 1991.

The greatest financial crisis in the history of the Vanderbilt Television News Archive came in 1991 at the outbreak of the Persian Gulf War. The leadership of the Archive made the decision to record as much news coverage of the war as possible. To explain the decision, we must briefly consider the Archive’s collection development policy.

The collection policy of the Vanderbilt Television News Archive has always had two foci. First, the Archive records the television news programming of the “big three” broadcast networks (ABC, CBS, and NBC) as well as a representative hour from cable news networks (CNN and Fox). Second, the Archive preserves breaking news stories and other televised events under a rubric it calls “Specials”. The precise definition of a “special” has elicited much discussion, but it includes events such as Presidential addresses, destructive weather events, terroristic attacks, and other major stories.

The commencement of the Persian Gulf War fell naturally into this category. Public interest in the war ran high. On January 18, 1991, Bill Carter reported in the *New York Times* that “President Bush’s address to the nation on the air strike against Iraq Wednesday night apparently attracted the largest audience in the history of American television” [7]. For the staff at the Television News Archive, there was another major reason for seeking to preserve all coverage of the Gulf War. At that time, the Gulf War was the most significant American military action since the Vietnam War. And the Vanderbilt Television News Archive started operating during the Tet Offensive in Vietnam and captured reporting about the war through the fall of Saigon in 1975. In the years since the war, soldiers and family members continue to query the archivists about distantly-recalled interviews and news footage featuring them or their loved ones.

The archivists imagined that something similar would take place in the wake of the Persian Gulf War. As they saw it, they were preserving the public record of the conflict while also serving its veterans by keeping records of their personal stories. As John Lynch, former director of the Archive, recalled in public hearings at the Library of Congress in 1996:

In less than two months we collected two thousand hours of materials. That is \$30,000 in tape, and an even larger cost in labor, since we had to spend about \$30,000 for a temporary position and at least \$10,000 in overtime costs. This decision to spend \$70,000 had to be made quickly. On January 16, 1991, the air war began and so did the taping. If we had waited even one day, we would have lost almost one hundred hours of network coverage [8].

To support the effort, the archivists envisioned a fundraising campaign that would provide veterans with personalized VHS copies of relevant news footage. As it happens, the ground conflict ended in less than two months and with light coalition casualties. As the recession of 1990–91 dragged on, the American public moved on to economic concerns. (“The economy, stupid” as James Carville famously insisted during the election campaign in 1992.) Meanwhile, the Vanderbilt Television News Archive had burned through most of its budget in the first two months of the year.

The *Chronicle of Higher Education* reported in 1993 that “while scholars have liked the system, it has proved too expensive for the university”.

This year, Vanderbilt administrators considered closing down the archive, which had accumulated a \$1.5-million debt from 1985 to 1992. They were dissuaded after the archive let nearly half of its employees go, revised its fee structure, and came up with a plan to distribute the index via the Internet [9].

This episode illustrates the financial limits of the Archive’s preservation program. From thirteen staff in 1991, the Archive dropped to eight in 1992 and then to five in 1995. Today, the staff continues to operate with five fulltime equivalent employees.

The goal of preserving the national evening news appeared straightforward, if daunting, at the beginning of the Archive. The three networks broadcast evening news shows running for thirty minutes. The arrival of cable news networks such as CNN in 1980 and Fox News in 1996 changed the calculus, forcing the Archive to preserve representative shows rather than all the major news programming.

The specter of loss haunts the Archive. While television viewership is on the decline, broadcast and cable news is still the single most important medium of information about current events for the American public. New networks also come into existence, providing another angle on the news events of the day. Think of *Al Jazeera America*, which lasted from 2013 to 2016. Or BNC (The Black News Channel), a cable news channel aimed at African-American audiences that launched on February 10, 2020 [10]. Beyond national news programming, consider local television programming across the United States, which remains, despite some significant exceptions (like the partnership between NBC 5/KXAS and the University of North Texas [11]), unpreserved.

As archivists, we cannot look to the past alone as what we choose to preserve depends on our anticipation about what will be significant in the future. As channels and networks expand, we have to make judgment calls about which will remain relevant in the future.

3. Technology

I mentioned at the opening of this paper that the Vanderbilt Television News Archives has approaching 1.2 million records in its database. These records provide, among other items of information, the date of

the original broadcast, the broadcast network, and the start and stop time of the particular segment. These database entries link out to media files that contain entire shows. We do not split the files into segments, but manually record their start and stop times in our metadata.

Owens notes that while having two copies of your digital media is essential and at least three is generally recommended, making copies is “ultimately useless if you don’t have a fully-articulated process for *managing* those copies” [12]. The philosophy of multiplying copies makes sense in environments with adequate metadata and storage capacity. But, when applied haphazardly, copies can threaten the vitality of a digital media archive. When I took on administrative responsibility for the Vanderbilt Television News Archive, I became overwhelmed by the number of copies of our media in our collection.

The previous generation of archivists decided to conserve the prior versions of the storage media even as they ported the collection to new formats. Given the longevity of the Archive, this means that we manage a sizable physical collection alongside our digital collection. As Jim Duran, the Director of the Archive explains, “The physical collection consists of thirty-five thousand 3/4 inch U-Matic tapes, duplicated on thirty-five thousand DVDs. Any failed or problematic transfers were saved resulting in a collection of a few hundred VHS tapes and less than fifty one- inch Ampex open reel video tape. Parts of the collection are also on sixty-two spinning disk hard drives (approximately 30 TB) with an unknown amount of duplication”. The question whether to conserve these prior physical media or to discard them after having ported their contents to new storage media remains unresolved. On the one hand, our storage space is not suitable for the longterm storage of film or electronic media. On the other, staff members continue to consult these physical archives when it turns out our master or access copies are corrupt or missing. The conversion of these prior formats means that we also maintain the machines necessary to retrieve their contents. Steve Davis, a longtime member of the staff, keeps this equipment up and running, but the suppliers for parts and service have become scarce.

From a preservation perspective, conserving these physical artifacts might constitute a form of “bad faith”. We hang on to the older formats because we are not sure that we have properly transferred them into digital formats. Practical grounds exist for this uncertainty. Staff had to retrieve videos from DVDs whenever patrons encountered problems with streaming media until Duran developed a software routine to check systematically for improperly-encoded media. As it happens, we are not alone in preserving our collection in both analog and digital format. As Leif Kramp notes in “The Complicated Preservation of the Television Heritage in a Digital Era”, television archivists “mistrust” digital preservation solutions, digitizing portions of their collections while maintaining their analog recordings [13]. This leaves archivists in the unenviable and expensive position of preserving analog and digital versions of the same material, creating “a considerable increase in the workload” [14].

Archivists will appreciate that the most significant threat to preserving our collection stems from our lack of metadata about the collection. We lack an inventory of the collection as a whole, meaning that we cannot answer questions about the number of copies of any given news show we have across all these media and formats. A strategic initiative at the Archive in 2020 is the creation of PBCore metadata for the entire collection [15]. As we generate metadata records, we will also include information about their so-called “instantiation” in both analog and digital media.

We ultimately want to reach the goal of having three preservation copies (MPEG-2/H.264 transport streams) of every news program in digital form: a copy on premises in Nashville, a copy on the Cloud (Amazon S3 Deep Glacier), and a copy at the Library of Congress. We can verify that we have met this goal through sharable, standardized, and comprehensible metadata. In particular, maintaining our preservation partnership with the Library of Congress requires that we agree on what we are both committed to preserving.

Backing up media and metadata in different locations is not enough. As James Luetkehoelter remarks, making plans to restore the data is also essential.

Remember that backup is only half the puzzle. Don't forget that you also must define a plan for a recovery process. Don't stint on the recovery half of the puzzle. Backups are usually done at leisure, without any undue stress. Recovery, on the other hand, occurs when you're mostly likely to be under stress, and consequently when you most need a well-defined and clear plan of action [16].

To my knowledge, the Vanderbilt Television News Archive has never experienced a major loss of data. But the threat of catastrophic system failure always lurks in the background. As an administrator, I feel like we are racing against time to describe our collection adequately as a prerequisite for restoring the collection in case of system failure. If a system failed, we would rely on our metadata as a guide for rebuilding the collection.

Here is a final quandary we face when developing the collection. As I mentioned, news channels are expanding as our staff continues to contract. While implementing new capture technologies, streamlining procedures, and relying more on student labor has helped us to bridge the gaps, we have put off making hard decisions about reducing our collection scope. But our ambitions outstrip our capacity, meaning that we are accruing technical debt.

Take the case of Fox News. The Archive began recording a daily hour of news programming from Fox News when the network launched in 1996. During the first three or four years, a generous grant from the Ford Foundation allowed us to describe the news according to our established standard. Since the end of the grant funding, the Archive has not been able to keep up, meaning that we record and preserve the media files without abstracting or indexing any of the contents. While you could say that we have preserved Fox News, we have failed to make it accessible.

4. Law

The biggest challenge that the founders of the Vanderbilt Television News Archive faced in 1968 was neither financial nor technical, rather it was a legal challenge. Was it legal for a third party not affiliated with any of the networks to record and preserve the evening television news? The technology for recording television programming was novel and posed a liability to the networks. At that time, the networks were not registering the copyright of the news shows. That television news merited registration was not self-evident as the copyright required content to be "fixed in a tangible medium of expression" to be copy-rightable. By taping television news programming off-air, was Vanderbilt tacitly claiming copyright on its recordings?

This is not the forum to review the disagreements and discussions between the Columbia Broadcasting Company (CBS) and Vanderbilt that led to the filing of a lawsuit against the Archive in 1973 [17]. Let it suffice to remark that a consequence of this episode was the legal consolidation of the Archive's loan program. On the model of a lending library, the Archive copies clips and shows from its collection and, on a cost-recovery basis, loans them to individuals for the period of a month. We send the loans of physical media, at first VHS tapes, then DVDs, and now USB thumb drives. At the end of the month, borrowers must return these devices to stay in good standing and to request further loans.

CBS threatened Vanderbilt again in 1994 as Vanderbilt prepared to put its index of news programming online. As happens in legal conflicts, the question at stake between Vanderbilt and the networks looks different from opposing perspectives. For the networks, the primary issue is control of the content.

Part of the argument is about money, says Don DeCesare, vice-president of operations at CBS News, but the real issue is copyright: 'There is only one side to the story: the material recorded is ours. It always has been and always will be', he says. 'If we put something on the air we stand behind it. But if someone else is duplicating it, editing it, and sending it out, we've been removed from that process' [18].

From Vanderbilt's side, the question is preservation of the public record. As Lyman Ray Patterson argued, prevention of access to the public record contravenes the fundamental *raison d'être* of copyright law.

These issues are not unlike the issues of press control in sixteenth century England. With its ephemeral transmissions, television eliminates not only the fundamental basis of copyright - the act of creation - but also the *quid pro quo* underlying the monopoly of copyright - the dissemination of copyrighted material in permanent form to the public. To give communications corporations a proprietary interest in public information and public domain materials would enhance their power to influence and shape the opinions of millions of people without imposing any means of making them accountable for the responsible exercise of this enormous power [19].

In March 2018, we held a Mellon-funded symposium on television news preservation to develop plans for a coordinated approach to television news preservation [20]. The copyright experts in our group noted that our activities benefited the networks by preserving broadcasts that otherwise would have been lost. After all, the point of copyright is to encourage the production of content, not to foster its destruction.

The legal landscape of television news preservation continues to grow more complicated as we must consider privacy safeguards like the General Data Protection Regulation (GDPR) [21] and the California Consumer Privacy Act (CCPA) [22] as well as the impact of legal opinions such as Fox News Network, LLC v. TVEyes, Inc. (2018) [23]. For librarians who are contemplating similar preservation efforts, the complexity of the legal environment may dissuade them from getting started.

In "Piracy Is the Future of Culture", Abigail De Kosnik argues that legal limitations and roadblocks to digital media preservation encourage so-called pirate archives.

While professional archivists have been stymied in their efforts to legally digitally copy and migrate cultural texts, pirate archivists have built up personal collections of digital cultural files and are sharing them freely online, allowing numerous exact copies of these files to be stored all over the world. Thus, pirate archivists have constructed what is essentially an alternative cultural preservation system... [24].

This phenomenon has already manifested itself in the field of television news archiving. The most prominent example is Marion Stokes, a former librarian who systematically recorded on forty thousand video cassettes the local and national broadcast news from the Philadelphia area [25]. Credit goes to Roger Macdonald and Brewster Kahle at the Internet Archive for preserving and making her videos accessible despite the inherent difficulties of digitizing and making sense of such an amorphous collection. Stokes numbers among extramural archivists who have preserved local audiovisual cultural heritage that might otherwise have become lost to history, as Jennifer VanderBurgh notes [26].

5. Future

The preservation of digital media takes place in conversation between the past and the future. As the archivists charged with responsibility for the collection in the present, we make critical decisions about continuity and discontinuity. As inheritors of the collection, we need to understand how and why the Archive came to have the shape it has today. As progenitors, we need to make sure that we prepare the Archive for its next generation of users.

The future of television news preservation is, of course, intrinsically linked with television news. Imagining a future without television is difficult. When the Marty McFly of 1985 arrived in 2015, he marveled at the digital Max Headroom of the future, but he still recognized the medium. Indeed, that vision of what television might be like in the future turns out not to be far from the case, as the Xinhua News Agency in China released its first digital anchors in 2019 [27]. Despite declines in viewership, television remains a robust medium and television news a persistent source of information for the American public.

The conversation between present and past affords an opportunity to reexamine whether our digital preservation efforts actually serve our mission. Are our current practices serving our mission of providing access to the public record of national news programming? In part, we are meeting that goal. Our loan system permits access to our entire collection and we also serve students and faculty at institutions of higher education through our sponsorship program. But what about informing the broader public? The price of the loans (\$27 per clip) effectively dissuades casual patrons as does the lack of any mobile interface. If preservation is ultimately about access, how can we make sure that the public enjoys access to our collection in a way that serves their information needs?

The future of television news is itself in flux. Will there still be national television news programming in five, ten, twenty years? When I was growing up, I could tell you the names of the anchors of the three major networks: Peter Jennings on ABC, Dan Rather on CBS, and Tom Brokaw on NBC. Will future generations have the same relation to television news? Or will they remember their favorite YouTubers as the primary sources of information? And who is preserving those news sources today?

References

- [1] P. Simpson, *Network Television News: Conviction, Controversy, and a Point of View*. Legacy Communications, Franklin, TN, 1995.
- [2] F.C. Schreibman, A succinct history of American television archives, *Film & History: An Interdisciplinary Journal of Film and Television Studies* **21**(2) (1991), 90.
- [3] H.J. Gans, *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*. Northwestern University Press, Evanston, 2004, p. 79, see also Braun, J.A., *This Program Is Brought to You By...: Distributing Television News Online*, New Haven, Yale University Press, 2015, p. 2.
- [4] P. Simpson, *Inquiry into Alleged Rigging of Television News Programs*. U.S. Government Printing Office, Getzville, New York, 1972, As Neil Postman and Steve Powers observe, this kind of "staging" persists in television news; see Neil Postman and Steve Powers, *How to Watch TV News*, Revised, Updated edition, New York, Penguin Books, Chapter Seven, 2008.
- [5] T. Owens, *The Theory and Craft of Digital Preservation*. Johns Hopkins University Press, Baltimore, 2018, p. 5.
- [6] E.M. Corrado and H.M. Sandy, *Digital Preservation for Libraries, Archives, and Museums*, second ed. Rowman & Littlefield, Lanham, MD, 2017, p. 6.
- [7] B. Carter, Giant TV audience for Bush's speech, *The New York Times* (1991).
- [8] J. Lynch, *The Current State of American Television and Video Preservation*. Washington, DC, 1996.
- [9] D.L. Wilson, Battle over a TV archive, *The Chronicle of Higher Education* (1993).
- [10] S. Andrew, Black News channel, the sole African American-led news network in the US, has arrived, *CNN* (2020).
- [11] M.D. Gieringer, A cooperative model for preserving historical television news context. in: *IFLA World Library and Information Congress News Media Satellite Meeting*. Columbus, OH, 2016.
- [12] T. Owens, *The Theory and Craft of Digital Preservation*, pp. 107–108.
- [13] L. Kramp, The complicated preservation of the television heritage in a digital era. in: *Information Storage: A Multidisciplinary Perspective*, C.S. Große and R. Drechsler (eds), Springer International Publishing, Cham, Switzerland, 2020, p. 224, doi:10.1007/978-3-030-19262-4_8.
- [14] Ibid.
- [15] On development and uses of the PBCore metadata standard, see N. Rubin, The PBCore metadata standard: A decade of evolution, *Journal of Digital Media Management* **1**(1) (2012), 55–68.
- [16] J. Luetkehoelter, *Pro SQL Server Disaster Recovery*. Apress, Berkeley, CA, 2008, p. 99.

- [17] M. Illson, Vanderbilt U. sued by C.B.S. on sales of Cronkite tapes, *New York Times* (1973).
- [18] B. Owen, Who owns old news? CBS takes on the Vanderbilt Archive, *Columbia Journalism Review* **33**(1) (1994), 18–20.
- [19] L.R. Patterson, Private copyright and public communication: Free speech endangered, *Vanderbilt Law Review* **28**(6) (1975), 1168.
- [20] C.B. Anderson and J. Duran, Sustaining television news archives, *Journal of Digital Media Management* **8**(1) (2019), 82–90.
- [21] See <https://gdpr-info.eu/>, last accessed June 15, 2020.
- [22] See <https://oag.ca.gov/privacy/ccpa>, last accessed June 15, 2020.
- [23] See <https://law.justia.com/cases/federal/appellate-courts/ca2/15-3885/15-3885-2018-02-27.html>, last accessed June 15, 2020.
- [24] A. De Kosnik, Piracy is the future of culture, *Third Text* (2019), 7. doi:10.1080/09528822.2019.1663687.
- [25] S. Kessler, The incredible story of Marion Stokes, Who single-handedly taped 35 years of TV news, *Fast Company* (2013).
- [26] J. VanderBurgh, Grounding TV's material heritage: Place-based projects that value or vilify amateur videocassette recordings of television, *VIEW Journal of European Television History and Culture* **8**(15) (2019), 59–78. doi:10.18146/2213-0969.2019.jethc165.
- [27] C. Aizhu, China's Xinhua presents news using robot news anchor, *Reuters* (2019), see <https://de.reuters.com/article/us-china-ai-broadcasting/chinas-xinhua-presents-news-using-robot-news-anchor-idUSKCNIQK0IU>, last accessed June 15, 2020.