Sustaining Television News for the Next Generation

Workshop Working Group: Technical

Summer 2018

On March 8-9, 2018 Vanderbilt University hosted a group of library and archive professionals to discuss the current challenges of television news archives. The event was sponsored by The Andrew W. Mellon Foundation. The group split into three working groups charged with specific issues. Each working group was given some questions to start the discussion, but they were also encouraged to raise other concerns or challenges they saw to their topic.

The technical working group met to discuss the various aspects of capturing and preserving television news. The conversations circled mostly around methods of producing computational metadata and gaps. They also discussed ways to improve textual indexing of collections. This working group was the smallest of the three groups with 5 to 6 members, including participants mostly from university libraries. Four of the participants had active television news archives and two had finite audiovisual collections that related to the conversation.

This working group first asked if this community of libraries and archives is on the cutting edge of recording, preserving, and making television news accessible or do we need to catch up? The question led to a discussion on the use of closed captioning for better discovery. The institutions represented by the group were at different levels of implementation of closed caption or SubRip subtitle (SRT) files for full-text searching. Those who were heavily using voice-to-text were doing so with machine learning computations or through a paid service for the text files. Participants commented about the accuracy of paid services and how much work was involved

in refining or cleaning up such resources. Over all, the use of closed caption files was accepted as a valuable source of text indexing.

The conversation continued about closed captioning and some challenges with relying on them for metadata. The group noted that political advertisements, for example, often lack closed captioning. Some other challenges to preserving the news included a discussion about the value in timestamping every word in a video archive, splitting audio by speakers and associating voice-to-text with corresponding speakers. The group agreed that researchers would appreciate queued results where a video database points searches directly to the search terms within a video clip.

As the conversation progressed, the group discovered that a couple institutions were building robust datasets of audiovisual descriptive information, like lists of recognizable faces or voices. The datasets are used for machine learning to automate the identification or individual's faces or voice. The group agreed that if these datasets were shared and collaboratively curated many television news archives could benefit.

The group also discussed some of the challenges with capturing and preserving non-traditional news formats like the platforms presented on social media. The biggest challenge to preserving these different types of news according to the consensus of the technical group is the lack of a consistent video format. The group agreed that time-coded metadata is key to any video archive and that with granular descriptions there are many possibilities for preserving audiovisual news.