

**Recommendation of the Federal Communications Commission (FCC) Disability Advisory
Committee (DAC) on Internet Protocol (IP) Closed Captioning Files**

Prepared by the IP Closed Captioning Files Working Group Working Group

Adopted by the Disability Advisory Committee on November 1, 2022

Introduction

This report, developed by the IP Closed Captioning Files Working Group (Working Group), addresses the Commission's request for the Committee's input on: (i) identifying issues and best practices regarding the transmittal and receipt of captioning data for full-length content viewed online, with the shared goal that captioning associated with such programming remains available regardless of the distribution method, and (ii) suggesting potential solutions to increase the amount of captioned programming online. As it relates to closed captioning, the Commission's role is defined in the Communications Act of 1934 as amended by the Telecommunications Act of 1996, the Communications and Video Accessibility Act (CVAA), the Television Decoder Circuitry Act of 1990, and other federal statutes, and the Commission's implementing regulations.

The DAC offers no opinion about (i) whether or to what extent the Commission has the requisite authority to act on the content of this Report or (ii) on any other aspect the scope of the Commission's authority. Nothing in this Report should be construed to concede any arguments by any member of the DAC related to the content of this Report, including in any Commission proceeding, related to the topics of this Report. This may include arguments

related to costs, benefits, burdens, the civil and human rights of people with disabilities, or other considerations.

Based on its interpretation of the Commission's request, the Committee focused its considerations primarily on issues attendant in the distribution of already-captioned programming and associated captioning data. Moreover, this Report is not intended to be an exhaustive discussion of issues related to the distribution of already-captioned programming, but instead provides a high-level overview of the current ecosystem and highlights technical, human, and organizational process challenges and opportunities to address them.

The Committee recommends that the Commission foster awareness of the Report among participants in the ecosystem, and work with participants to address the challenges and pursue the opportunities identified in the Report.

Report

I. The Closed Captioning Ecosystem

The captioning ecosystem for video programming online is large and complex. This section discusses the process by which programming and associated captions traverse the ecosystem and notes complications that may arise at various points in the process.

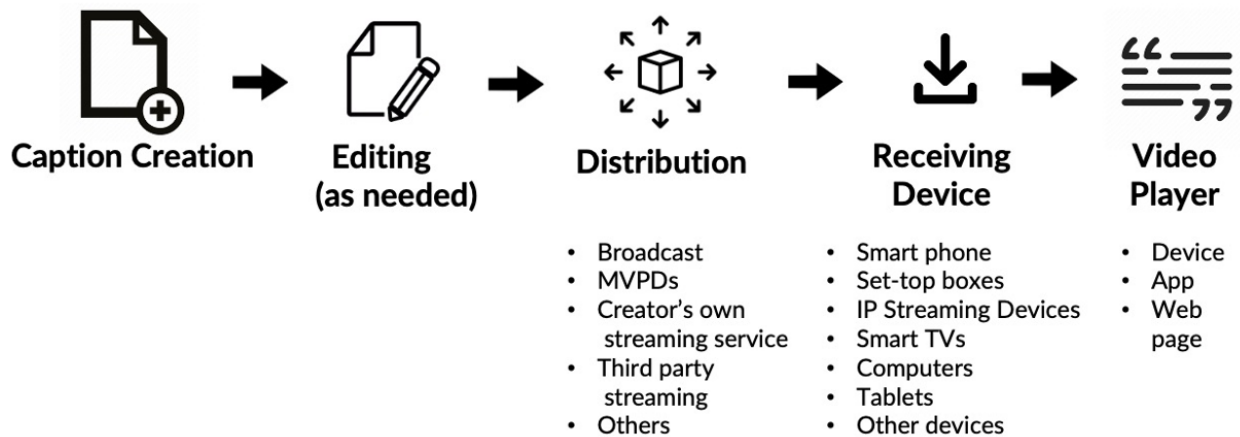


Figure 1—High-Level Captioning Workflow

Figure 1 above shows in a simplified manner the steps involved in captioning full length IP-delivered programming. We explain the process in additional detail below.

A. Caption Creation

Captions are initially created by a content creator, owner, or provider, either internally or through a vendor. They are sidecar components to a video program and may be delivered as a file embedded¹ in the media container,² or may be streamed alongside the video, such as for most online program delivery, including live streams. After creation, the

¹ Embedded captions mean that the video signal also includes the caption information

² Media containers format how different multimedia data elements (streams) and metadata coexist in files. A media container is used to display media (video, audio, or other data) with a caption.

captions may be converted to various formats—e.g., SRT³, VTT⁴, or SCC⁵—as required by the content distributor or to suit the method (or platform) of distribution—e.g., broadcast, cable or other multichannel video programming distributors (MVPDs), streaming—and technology associated with a particular delivery platform. As an individual program may be distributed in multiple ways, the captions may need to be converted to multiple different formats or specifications.

B. Editing Caption Data

Captions may be edited at various points in the workflow as appropriate for the diverse uses of and distribution mechanisms for content. For instance, pre-recorded broadcast/cable programming may be reformatted for timing and commercial insertion (which can greatly vary among services), then played out from a national or regional source, where it is encoded in multiple transmission formats (e.g., standard definition, high definition, ABR⁶). Each format may have unique caption encapsulation methods. In

³ SubRip Text (SRT) is one of the most common types of raw closed caption formats, typically ordered after a video has been produced (i.e., the post-production process). The captions can be uploaded to the platform where the captions will be hosted as a sidecar file separate from the video. <https://en.wikipedia.org/wiki/SubRip>.

⁴ Video Text Track (VTT or WebVTT) is a popular subtitle and caption format. <https://en.wikipedia.org/wiki/WebVTT>.

⁵ Scenarist Closed Caption (SCC), which are commonly used with broadcast and web video, as well as DVDs and VHS. https://en.wikipedia.org/wiki/Closed_captioning.

⁶ Adaptive bitrate (ABR) streaming is a technique for dynamically adjusting the compression level and video quality of a stream to match bandwidth availability.

contrast, for IP-delivered on-demand content, the supplier may translate/convert the content to multiple delivery formats which may include caption format translation.

There are several types of edits that can impact the program duration and the accuracy or synchronicity of the associated captions. These types include technical format edits, recuts of a show or film (e.g., Directors' Cuts), edits for length, video clips, as well as other modifications and insertions (e.g., service identifiers that indicate the origin of the content, advertising interstitials⁷).

C. Distribution

Once converted—and depending on the distributor's agreement with the content provider or lack thereof—the content provider may deliver both the video and, time-coded caption data to the content distributor. The data may take a variety of delivery forms—e.g., embedded captions or sidecar files. If a content distributor offers programming to consumers in more than one way, it may receive more than one video and/or caption format depending on the technical needs of its distribution platform(s).

For example, a video may feature a certain type of embedded captioning in addition to an external associated caption data. Distributors can select among caption formats or convert the captions into additional formats as needed. The content distributor may also perform additional work on the caption data, such as corrections, or conversion to additional human languages. The distributor will then make the media, including captions,

⁷ Advertising interstitials are ads placed between content, typically at natural breaks or transition points in a program, such as between scenes.

available for transmission over its distribution platform to the end-user's receiving devices. Captions will be rendered through the device or an application.

D. Receiving Device

One issue with receiving devices is that captioning formats associated with an older device might produce an inconsistent experience compared to a recently manufactured device. Indeed, certain arcane captioning format specifications (like .cap) are outdated and may not translate well to newer technology. Additionally, many devices offer their own proprietary player environment, which can result in different viewing experiences as discussed below.

E. Video Player

Depending on the digital platform used by the viewer, either the content provider or a third party displays digital programming through a software or firmware⁸ video player with captioning functionality. When captions are embedded as data in the video stream, the software or firmware player renders and overlays the caption text on the display screen. For pre-captioned programming, the existence of multiple specifications among software development kit⁹ (SDK) manufacturers means that programmers need to deliver multiple captioning format types for digital delivery of full-length content.

⁸ Firmware refers to software that has been permanently installed in a machine, device, or microchip, usually by the manufacturer, without which the device will not work. Firmware typically operates in the background without much interaction with users.

⁹ A software development kit (SDK) is a set of tools used for developing applications that is provided by hardware and software providers, usually including application programming interfaces (APIs), sample code and related documentation.

F. Differences Among Cable, Broadcast, and Online/Streaming Content Workflows

To contextualize some of the above-discussed complexities and challenges for online full-length captioned programming originating from multiple sources, consider, for example the following file-based captioning workflows specific to certain broadcast/cable programming and online/streamed content:

- *Pre-produced (or, episodic) programs¹⁰ with a broadcast or traditional cable layout.* The broadcast or cable network's studio sends recorded video footage to a captioning vendor— often a third party—which in turn, provides time-coded captions, typically within eight to 72 hours. For some types of distribution, caption “sidecars” may be embedded by the distributor in the audio/video content prior to distributing to digital platforms.
- *Live-produced programs, including news and sports with captioning.* Due to the short time for delivery of data from camera to digital player, there is less preparation time and human-generated captions are created in parallel with content transmission and added at the playout encoder¹¹ at the national/regional/or local source.
- *On-demand programs.* Content made available on-demand via a website, app, set top box or streaming service is received by the content aggregator, in either

¹⁰ These are often scripted programs that go through a post-production process and can have more refined and accurate captions relative to live programs.

¹¹ Playout encoders convert data from one format to another, such as converting media assets to streamable data. Encoded files must be compatible with end user devices' decoders.

separate components or as a complete encoded single file. Alternatively, certain content might utilize automatic speech recognition¹² (ASR) technology as a primary captioning source with or without human oversight, depending on the platform.

Given the complexities detailed above, some content creators and distributors have implemented quality control processes to manage the technical syntax of the captioning data stream and facilitate the exchange of captions within the ecosystem. This may include practices such as:

- Developing policies and procedures to test pass-through of captioning throughout their own distribution chains;
- Developing policies and procedures to help ensure the consistency of captioning availability across video platforms and devices; and
- Using automated technologies and human oversight to monitor, detect, and correct captioning issues in catalogs of programming.

While multiple stakeholders typically engage in quality control processes, the technology used to perform such reviews varies greatly.

II. Potential Challenges in the Captioning Ecosystem

Because many parties are involved in delivering captions to the consumer, both technical and human organizational process challenges may arise.

¹² Automatic speech recognition (ASR) is the use of machine learning or artificial intelligence to process human speech into text.

A. Potential Technical Challenges

Some technical challenges in the captioning ecosystem include:

- *Caption corruption.* Captions may unintentionally become corrupted at some point in the workflow.
- *Timecode shifts.* These may arise when a service or content owner inadvertently changes the time code of the content, by adding interstitial content like advertising or service bumpers. Other causes may be frame rate conversion problems caused by bugs in the conversion software.
- *Captions for certain older programming.* Older programming may have captions dependent on older technologies, preventing distribution and/or display.
- *Software bugs.* There may be software bugs in at various points in the chain—server bugs, client/player parsing or rendering bugs, ingest bugs, etc.
- *Editing tools.* Re-editing a video that was already captioned requires access to the program master and the original segmented caption master, as well as an editing suite that is able to keep the captions in sync during the editing process. The participant in the ecosystem may or may not have access to these caption and program masters or tools, so captions may need to be recreated if a program's content or timing are changed.

B. Potential Human and Organizational Process Challenges

In addition to technical challenges, some human and organizational process challenges include:

- *Caption is missing through the distribution chain.* Because of the different technical formats and various procedures for sharing captions among vendors and other third parties in the production and distribution chain—e.g., content creators, license holders, captioning vendors, content distributors, etc.—captions may be lost through the distribution chain.
- *Captioning is available but may not be usable by end user.* This may include situations in which captions are created at various parts of the distribution chain (e.g., use of ASR at or near the video player stage described above).
- *Lack of expertise or awareness among small content creators.* Some content creators may lack the expertise to caption programming or the awareness to do so.
- *Lack of awareness among app developers.* Some app developers may create apps with playback interfaces that do not support captions or that lack unified specifications. Some app developers may also use playback methods for advertising that impede placement of captions.
- *Recut or edited video.* As mentioned above, some editing tools exist that enable video and accompanying captions to be cut so that captions continue to match the video, but those tools rely on the availability of program and caption masters. Even with the availability of such tools, content creators may nonetheless need to use an outside vendor to ensure captions match edited video. For example, when a caption sidecar file is created, it is timed to match the corresponding video. If the video is recut or edited in a manner that changes the timing, the captioning sidecar file will no longer match, and the captions will need to be adjusted to

match the video, as discussed above either by using editing tools or delegating to an outside captioning vendor.

- *Lack of awareness among those authoring content creation, ownership, and delivery contracts.* Given the infancy of the streaming video ecosystem relative to traditional television, contracts may not exist in some instances, or where they do exist, they may not cover captioning, with the exception of large provider and delivery contracts that may define expectations and reconciliation steps in the event a content owner does not have or cannot deliver captions.

III. Potential Opportunities in the Closed Captioning Ecosystem

Some opportunities for participants in the captioning ecosystem to address the challenges in the previous section include:

- Developing internal policies and procedures¹³ to:
 - Test pass-through of captioning throughout their own distribution chains;
 - Help ensure the consistency of captioning availability across video platforms and devices; and
 - Include captioning delivery requirements in licensing and distribution agreements;

¹³ Although we acknowledge that aspects of these three policies and procedures may mirror activities that the Commission requires of some entities, the DAC takes no position on the scope of the Commission's rules, as noted above.

- Where quality control issues are observed, offering separate caption delivery to avoid re-exporting full video assets, particularly for live-produced programming;
- Developing opportunities for content owners, captioning vendors, and programming distributors to support the secure exchange and sharing of captioning files amongst themselves to avoid duplication of captioning work;

Some opportunities for the Commission include:

- Encouraging app stores, captioning vendors, distribution platforms, and other organizations to promote education, awareness, and documentation of development and use of accessible video technologies, APIs¹⁴, and development practices to help ensure that app developers incorporate captioning support;
- Encouraging efforts by developers, vendors, and providers of professional and amateur video editing, production, and distribution tools to ensure that tools are available to creators to promote the availability and utility of captions in their videos. In order to support content creators at all levels of sophistication, these efforts could include standalone applications and tools built into social media and video hosting platforms, tools that are capable of preserving captions as edits are made to already-captioned videos, and tools that resolve caption synchronization issues using machine learning and A.I.; and

¹⁴ Application program interface (API) defines how two apps communicate with each other and is often used to automate the process of adding captions to a video online.

- Educating video creators at all levels about the importance of treating captioning as an integral part of the creative process.