

Subtitling & captioning challenges in the modern workflow



An evolving landscape

Television is evolving fast – from multi-format delivery over multiple platforms to the interactivity of the network. There is a necessity to automate and integrate production and delivery workflows more seamlessly than ever before and repurposing content with minimal effort.

Although there are great opportunities presented by the proliferation of platforms, it remains essential to meet the challenges presented within these arenas to ensure ancillary data (such as interactivity, captions, subtitles, parental controls, etc) is maintained.

This paper focuses on captioning and subtitling, though many of the techniques employed to enable these technologies are applicable to other forms of ancillary data too.

Why do we caption & subtitle content?

First, let's consider why captions and subtitles are played out at all!

There are generally two reasons for captioning and/or subtitling content, firstly for accessibility – both for the hearing impaired and for when audio is unavailable (e.g. within public spaces such as at airports, gymnasiums, etc) – and secondly for language translation. A welcome side effect to providing captions or subtitles is that this metadata allows for searching of the content.

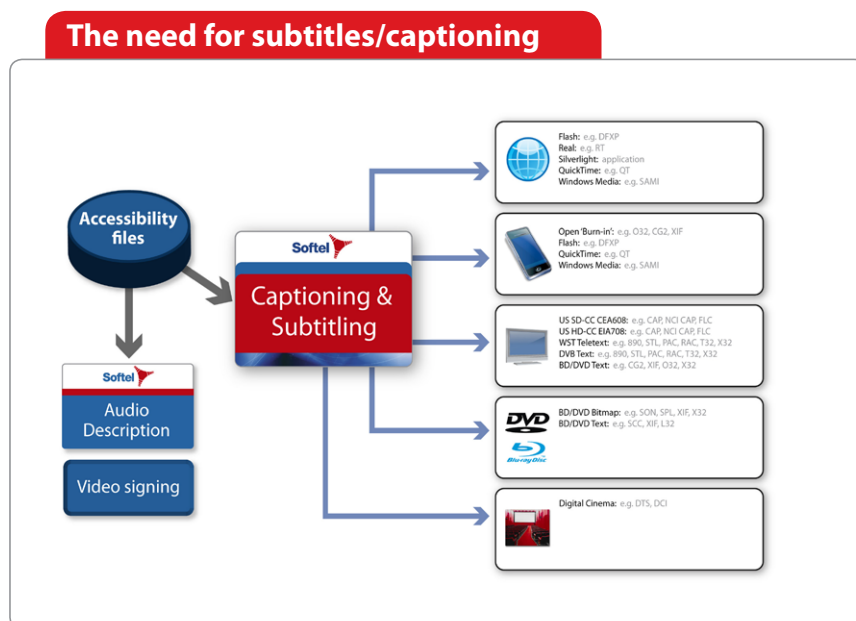


FIGURE 1: Deployment of Captioning & Subtitling to a Multitude of Platforms and Standards.

In the main, captioning in North America is driven by mandated requirements for content to be captioned when broadcast, and bills are passing through Congress to extend this requirement to Web content too. But as well as providing captions for the greater good and for accessibility there are also commercial reasons for offering the service. In North America, more than 10% of the viewership use captions due to hearing impairment. Equally, and as noted above, captions allow for content to be accessible in public spaces. Captions provide this accessibility not just to the main program content, but also to advertisements too, widening the demographic served by the advertiser.

A significant rise in international feeds which originate and are served from North America means that demand for subtitling is at an all time high. Subtitling and captioning are very similar and the terminology is often mixed and muddled. Alas in different parts of the world, the phrases mean opposite things. However, for the most part, captions usually refer to NTSC/ATSC standards such as CEA-608 and CEA-708, while subtitling either refers to an open (non-switchable) presentation, or to standards for the rest of the world, such as WST Teletext and DVB (bitmap or textual).

Reaching global multi-language audiences presents a challenge to broadcasters, but commercial strategies and competitive forces are driving the need to meet this challenge. While for some regions it makes sense in terms of workflow and cost to dub certain content, dubbing is only part of the solution, as, for most broadcasters' output, dubbing is not a feasible option for a number of reasons. The cost of dubbing dialogue means that while it may be appropriate for premium content — major movies being the prime example — there is not the budget to dub everyday drama, documentary and public information content in the same way.

The dubbing cycle is also time-consuming, ruling it out for a large portion of content. While the cost and time required for dubbing into one language means that broadcasters can reach majority-language sections of their potential audience, this may still leave many viewers in a multi-lingual region watching content dubbed into a language that is not their own. In many regions dubbing is also not culturally accepted and, obviously, still does not provide access for hearing impaired viewers.

The alternative — subtitling — provides a number of advantages that make it a very attractive solution for reaching the greatest number of viewers: the reduced cost of subtitling when compared with dubbing makes it more economically viable for a greater range of content; efficient state-of-the-art subtitling workflows mean content can be subtitled in much less time; and it is easier and more practical to offer subtitles, especially for content that is live-to-air or is has completed production shortly before time-of-air.

Through subtitling broadcasters can reach many different language groups within their audience, ensuring that viewers can fully appreciate the meaning of content, whether it is entertainment, information, promotion, or even advertising.

Dependent on the format of subtitling used, it is possible for multiple languages of translation to be delivered to the viewer, and for the viewer to choose their preferred language via their remote. In formats such as WST Teletext and DVB, it is also possible to switch on and off subtitles via the remote control.

So as well as the necessity to comply with legislation in the target international location, subtitles provide an inexpensive way to open up content to the wider local and international audiences through language translation.

Subtitles and captions also provide instant metadata for the video asset. This data helps add value to the video asset by increasing its “searchability” which, amongst other things, aids in repurposing.

Evolving workflows

Like any aspect of broadcasting, captioning and subtitling can be a slow, labor-intensive process if done with outmoded methods. So the key to using captioning and subtitling extensively and effectively is an efficient workflow that allows broadcasters to shorten the creation and playout cycle and keep costs at a manageable level.

All broadcasters have different requirements and processes and, as such, need individually tailored workflow solutions.

The captioning component needs to be integrated into the broadcasters overall solution and ideally considered during the initial design of the system. With a goal of supporting a multitude of output video formats, the focus of the broadcasters' systems are shifting away from the traditional production systems and transmission chain, towards its Digital Asset Management System (DAMS).

As broadcasters, networks, and content originators switch to file-based operations, the traditional captioning workflow has ceased to be a viable way of handling the growing requirement for captioning, especially as much content is now delivered across multiple platforms and in multiple formats.

The modern captioning workflow is extremely streamlined, but there remain several important steps to getting captions and subtitles to viewers beginning with the creation process and then either tape-based re-mastering, ingesting to video server, or time-of-air transmission. Below we discuss these steps.

Caption & subtitle creation

Many broadcasters have in-house caption creation expertise, though often these departments focus on quality control and repurposing efforts. Indeed, a significant number of content providers choose to outsource the creation of their captions & subtitles to specialist agencies. An ever increasing number of these agencies have sophisticated platforms to make the process of subtitle creation as efficient as possible.

For example, most use creation technologies which fully support non-linear video – thereby allowing content providers to digitally send encrypted and/or watermarked video clips. This saves costly couriering of tapes, as well as the time to make the tapes and the cost of the physical tapes themselves. The best creation systems allow for far greater captioner throughput through productivity enhancing technologies such as shot change detection, semi-automated timecoding functions, advanced reading speed algorithms and so on.

Moreover, the employment of a wide network of captioners makes expedited delivery ever more possible. Often these captioners are located around the globe in order to maximize time zone benefits. In fact a very substantial worldwide network of professional freelance captioners and subtitlers has emerged, each equipped with the latest creation workstations and using high-speed broadband links to receive secured video clips and job instructions.

Some modern caption authoring systems are Unicode capable, meaning they can create and repurpose

subtitles in almost any language, and many now fully cater for HD and digital cinema.

Finally, creation is sped-up through agencies sending “proofs” back to their clients electronically. This is achieved by generating an all-digital “approval”, which has subtitles overlaid to video so that the client can quickly and easily assess placement, timing, font choice, and other factors.

A note on file formats

It would not be right to discuss captioning and subtitling without mentioning the massive array of file formats which we must deal with within the space. Although there are several newer formats with the promise of being the new all encompassing standard for the future, the number of legacy files in the field means any system must cope well with both the import, repurposing, and export of a huge number of formats. Broadcasters and content producers have simply invested too much to discard legacy files. Some of the most influential file formats were shown in Figure 1 previously. The vast majority are proprietary and often very guarded, vendor-specific formats.

Great care must be taken when cross-converting formats to ensure the preservation of metadata and other essential information about the captions or subtitles.

Delivery to the viewer – “binding”

Once the created caption or subtitle file is signed off for delivery, it must then be “bound” to the content such that it is presented to the viewer when they watch the programming.

We tend to consider this binding as attached to one of three periods of time:

- Early binding – The pre-prepared file is matched to the programming well ahead of transmission;
- Late binding – Similar to early binding, but occurs near to air time and is only possible thanks to faster-than-real-time encoding techniques;
- Live binding – For either truly live content or for pre-prepared content which only becomes available very close to airing thereby eliminating the possibility of pre-preparing captions, a live bind is the technique employed.

Historically pre-prepared content was early-bound by creating a sub-master tape which would have the captions encoded into the VBI space on the tape by inserting into baseband video. Although this is still possible, it has become a somewhat outmoded method due to it being so labor intensive and slow and is therefore being phased out by almost all.

Instead, files are now either sent for time-of-air transmission (a live bind), or are transcoded into a file-based video asset (during early or late binding).

Time-of-air transmission generally involves systems which integrate into the automated workflow of a master-control facility with the subtitle playback

system approving files for playout in advance of airing and then airing the correct file at the right time automatically either with or without timecode.

The time-of-air system can also be used as a gatekeeper for real-time captioning whereby the system authenticates the captioner and their work slot prior to allowing pass through to air.

This system of checks is useful given the distributed and freelance nature of real-time captioning. Freely providing modem dial-in details could - and has - led to security issues with unsuitable caption content reaching the viewer without the knowledge or approval of the broadcaster.

Additionally, the time-of-air system extends the control of the automation system over the live captioning by switching the caption data source to the in-line caption encoder based on the automation schedule, removing the dependency on the live captioner to remember to do the right thing.

Increasingly the workflow of choice is a hybrid of ingest and time-of-air methods. This results in ingest to video servers whenever possible and allows for time-of-air playout as appropriate. The time-of-air caption system receives the playlist information directly from the automation system. Issues such as missing files, missing timecode, media lacking encoded captions or missing metadata can be reflected directly back to the automation system such that MCR staff are able

to take remedial action as soon as problems are spotted and well in advance of on-air problems. Where the automation playlist indicates that the video asset already contains caption data, the time-of-air caption system can check that it is complete, flagging any errors appropriately.

The time-of-air caption system can also provide interfaces to other ancillary data signals and XDS information such as wide-screen signaling, vChip parental controls, Broadcast Flag information, DRM controls such as CGMS-A data, Digital Program Insertion (DPI) data, etc.

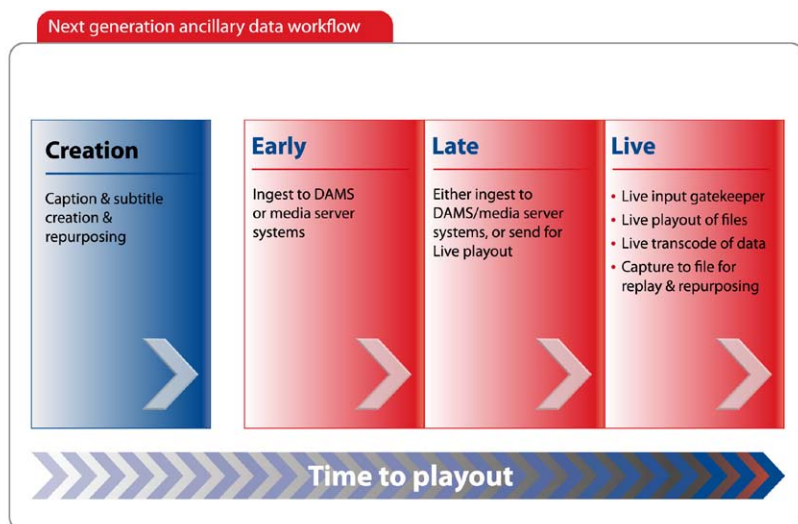


FIGURE 3: Binding of Captions & Subtitles.

AREA OF WORKFLOW				
Preparation		Binding		
Task	Caption & subtitle Creation & Repurposing	Early Binding	Late Binding	Live Binding
	Create caption & subtitle files in various formats ready for ingest, distribution or transmission	Optionally either create tapes or ingest caption & subtitle data to asset management or video server	Determine whether there is time to late bind (per early bind) or if live binding is necessary	Encode/ingest or inject captions & subtitles at time-of-'air'/'stream'; input security; capture for re-purposing

FIGURE 4: Areas of Workflow.

Video server technologies continue to roll-out at a rapid pace with companies such as Grass Valley, Harris, and Omneon leading the charge. Again, it is important to ingest ancillary data (such as subtitles) onto these servers. Systems are therefore required which will repurpose content ready to be ingested, with Application Programming Interfaces (APIs) developed to assist in the exchange of data around the modern workflow.

Moreover, Material eXchange Format (MXF) provides an open standard wrapper for broadcast media including ancillary data. Many equipment vendors have signed up to support MXF thereby enabling seamless workflows. MXF is an extremely flexible specification and most broadcasters who use the standard design a profile or "shim" to remove some of the ambiguity from this comprehensive and far reaching specification. This profile should also include the definition of how the captioning data should be supported.

Adapting to moving targets through "transcoding"

Increasingly, broadcasters are required to generate multiple versions of the same content for alternative and, in some cases, non-broadcast distribution such as online video. The myriad of distribution methods employ a variety of captioning and subtitling technologies which need to be supported.

By using the mezzanine video format described above and including the über subtitle file, the caption data component can be transcoded appropriately at the same time as the video ensuring the same quality, or better, of subtitles as for the broadcast version. The same also applies for situations where the 'lowest common denominator' mezzanine route has been followed, which while it may not provide better presentation on playout, the other benefits of easier repurposing and greater efficiency through predictable output still apply.

In addition to supporting different output distribution formats, it is often required to support "reversion" outputs where the video asset is manipulated time-wise or split in different program segments. This "reversioning" process often occurs within a Non-Linear Video Editor (NLE) and can effectively destroy the captioning data. Modern transcoding solutions can circumvent this issue by using the Edit Decision List (EDL) from the NLE thereby bridging the caption data from the original to the final version.

Conclusions

Captions and subtitles are an important piece of the broadcast jigsaw, but new thinking and new workflows are essential to keep pace with the roll-out of new platforms, global distribution and evolving standards. More automated and seamless workflows are not just a nice to have, but are increasingly an essential in order to allow for cost effective yet robust delivery to the widest content audience possible.

Appendix 1 – Capabilities within various standards

	SD CC (CEA-608)	Digital CC (CEA-708)	WST Teletext	DVB Text	DVB Bitmap	Open
Live supported	Yes	Yes	Yes	Yes	Yes	Yes
Customizable Aesthetics	-	Yes	-	-	Yes	Yes
Basic Languages	Yes	Yes	Yes	Yes	Yes	Yes
Unicode Languages	-	Yes	-	-	Yes	Yes
Switchable by Viewer	Yes	Yes	Yes	Yes	Yes	-
Open	-	-	-	-	-	Yes
Capture for re-purposing	Yes	Yes	Yes	Yes	-	-

Figure 1: Capabilities Matrix – Broadcast

	Flash	Quicktime	Windows Media	Silverlight	Mobile/ Portable Device	Real
Live natively supported	Yes	No (only via open)	Yes	No	Yes*	No
Simulcast	Yes	No	Yes	No	Yes*	No
Customizable Aesthetics	Yes (via DFXP)	Yes (via track file descriptors)	Yes	Yes	Yes*	Yes
Basic Languages	Yes (via DFXP)	Yes	Yes	Yes	Yes*	Yes
Unicode Languages	Yes	Yes	Yes	Yes	Yes*	Choice of coding (e.g. Japanese – JIS)
Switchable by Viewer	Yes	Yes	Yes	Yes	Not usually (separate stream needed)	Yes
Open	Yes	Yes	Yes	Yes	Yes	Yes
Capture for re-purposing	Yes	No	Yes	No	No	No

Figure 2: Capabilities Matrix for New Media.

* Please note that this may vary depending on your mobile platform

