Stream and File Formats—Where Are We Now?

By Bob Edge

Moving broadcast material in the compressed domain as data files with file transfer protocols can maintain image quality, reduce operating costs, and enhance operational capabilities. File formats have evolved over several years: some have been standardized, others are being held as proprietary technology, and new formats are being introduced. Digital Moving Picture Exchange, the first to be standardized by SMPTE, is used for uncompressed material transfers by telecine and synthetic image rendering systems. The next, General Exchange Format, is intended for news, sports, and on-air operations. The AAF Association has released an audio/video editing tools set designed for post-production and rich-editing applications. The Pro-MPEG Forum is working with SMPTE to develop a new file format, the Material Exchange Format, which will support a broad range of applications. Broadcasters must examine their requirements and choose a file or streaming format that has the features that meet their needs.

Content goes through many steps as it moves from a camera to a television receiver and an archive. Realtime analog or digital streams are used for many material transfers today, but they cannot guarantee perfect transfers. Bit-perfect transfers using file formats can eliminate repeated quality assurance steps, thereby improving a facility's productivity.

Each application presents unique requirements. For example, news has timely presentation as its highest priority, sometimes at the expense of quality; commercials are presented at the highest quality; sports production is a mix of video from a live event and material that is produced in advance. The pass-through or time shifting of network programs and the presentation of promos and public service announcements are part of every operation. As broadcasters start providing internet content, material is needed in one or more internet streaming formats.

Post-production is a different world. A large collection of new and archival material is combined in an artistic project. Artistic goals, project deadlines, and cost constraints drive the work. Increasingly complex effects and more descriptive metadata create evolving requirements that must be met.

Presented at the 143rd SMPTE Technical Conference (paper no. 143-15), in New York City, November 4-7, 2001. Bob Edge is with Grass Valley Group, Beaverton, OR 97006 An unedited version of this paper appears in *Pixels, Packets, Processing, and Infrastructure,* SMPTE, 2001.

As the cost of computing equipment falls and computer, storage, and network performance and capacity improve, it becomes practical for computer-based devices to store and process video streams. One result of this evolution is the nonlinear editor (NLE). Another is the video server or digital disk recorder.

With the increased acceptance of video servers and NLEs, the need to move material between these devices as compressed streams also grows. Compressed material can be transferred as files on data networks, data tape, or removable disks. When content is moved in the compressed domain as files, metadata management is simplified and no generation (decompression and recompression) quality losses are introduced.

Existing formats serve many users' needs in today's operations. Each stream or file format has its strengths, and matching these features to operational requirements is key. In some cases, new formats offer advantages. In others, the cost of change is a significant barrier.

Interoperability with File and Stream Formats

Fundamental conflicts make it difficult to design a single format that meets all requirements. For live events, low latency is important and basic stream formats such as conventional video, DV, and I-frame-only MPEG meet this requirement. A post-production house can gain significant benefits from a file format that can save work in process, since subsequent re-editing and repurposing are less difficult. Formats for on-air or station operations are designed to offer moderate latency and editing features required for last-minute changes. Voice-overs, cuts-only edits, and audio fade in/out are key features.

For example, if a file is stored as a collection of audio or video tracks, an NLE can work with it quickly and effectively, but a compile step is required to turn independent tracks into a low-latency time-multiplexed audio/video stream ready for playout. This step may include rendering

of dissolves and other complex transitions. Filtering the metadata is another part of the compile process. The compile step converts the format from an easy-to-edit structure to a stream that is optimized for operations and playout.

Another example is the need for a scheme to enable stream cutting or retrieval of a portion of a file from an archive. If frames have a fixed length, as in DV compression, locating the desired frames is comparatively simple. For compression systems with variable bit rates such as long-GoP MPEG, finding a specific frame in a stream may be difficult and will require long pre-roll times. Reading all the data sequentially does not exploit high-speed search capabilities of most data tape drives, which can search at 60 times their normal read/write speeds.

The type of frame lookup table and its structure depends on the compression scheme. A long GoP MPEG stream must be pre-rolled from an I-frame preceding the cut. For DV compression, a frame lookup table is unnecessary since the compressed frames are a fixed size. Future compression technologies could require yet another frame lookup scheme.

In some file interchange formats, a complete frame lookup table is implemented; in others, a partial one is used to enable retrieval of a segment from an archived file. Another class of formats does not include a frame lookup table at all; a table is created by the receiving system if needed. Conflicting application requirements result in multiple frame lookup table designs.

Applications Requirements for File Formats

At first, it would appear that a single file format would simplify a user's operation; but the diversity of requirements makes this difficult. An overview of parts of a broadcaster's material chain and the file format implications follows.

Acquisition and Contribution for Production

News and live events, especially interviews between geographically separated locations, require low latency. The results of adding file format latencies for interviews produces unacceptable hesitation in the verbal exchanges between the anchor and the remote talent. In a news environment, there will always be some instances where only low-bandwidth connections are available. In some cases, getting the news on air is more important than quality-loss or high latency. In these situations, news operations tend to use what they can get.

Sports programs are live events with pre-produced segments inserted during the event. Production requirements differ for sports and news, and these differences need to be taken into consideration. First is the ability to set up substantial facilities for a sports event. While sports events are planned, news events are not. The need to control latency and perform complex realtime effects makes traditional uncompressed audio and video distribution systems the best solution for live events. Managing pre-produced segments can be simplified with appropriate file formats.

Post-Production

Post-production material must be of very high quality. In addition, a rich collection of metadata is part of most post-production systems. A typical system stores the metadata and an edit decision list (EDL) in a database. This database may be implemented in a way that encapsulates the material, or it may have references to the video and audio tracks that are stored as separate files, or as material on tape or film.

Stream formats are used to store material in these systems. In addition, a rich data model is needed to support the easy-to-use and responsive interfaces editors require. The richness of the data model and its ability to describe the various effects and processing that a stream will undergo establishes a system's capability.

When a project is in post-production, it may be desirable to save a copy of the work in process. This allows easy continuation of an editing session after the program is sent out for approval. Using a file format that keeps a work-in-process version of a final program also simplifies production of derived works.

A post-production composition is compiled or conformed to convert to a streaming format. This step eliminates the EDL and most of the production and editing metadata. This is the handoff from the post-production and mastering facility to others in the content distribution chain. Only metadata essential to the finished product is carried forward with the final master. A post-production facility may archive both the final conformed production and the edited master data model for future projects.

On-Air Operations

The day-to-day operation of a broadcast facility presents a unique set of file format requirements. The rich editing and material production metadata used for production and post-production is not needed in the finished material. However, metadata about ownership, content usage, and related information is useful.

A broadcaster needs access to program content with latencies of a few seconds. A format that supports voice-overs is needed to accommodate last-minute changes. Cuts-only video edits with audio fade in/out are useful.

Using file formats with bit-perfect file transfers on high-speed data networks can improve operational efficiency. Material can be moved to and from archives or staged between central storage systems to playout devices easily and without quality loss.

Archives

Each facility imposes different requirements on its archives. A station may have archival requirements mandated by government agencies, as well as other policies that are established to meet business needs. In either case, the critical requirements are stability and simplicity in the storage format. If the format is not stable, decoders will not be available to support access to the material. The more com-

plex a format, the more likely that different manufacturers will inadvertently implement it slightly differently leading to long-term interoperability problems.

Metadata and File Formats

File formats designed for post-production and rich editing systems use metadata to store the edit decisions, editing history, material ownership, and more. Acquisition, on-air operations, and emission formats utilize less metadata.

"Sports programs are live events with pre-produced segments inserted during the event. Production requirements differ for sports and news...."

Management of metadata during the production process is growing in importance. Metadata for shot and take information, shooting location, timing, and links to the script or storyboard are critical to improving production efficiency. In news acquisition, being able to capture "who, what, when, where, and why" information during a shoot can improve access to this source material in later phases of production or for subsequent related stories. It is more difficult to accomplish this in a breaking news environment. Binding metadata with the audio and video content improves the efficiency of later post-production and archiving.

The metadata requirements in a sports material chain are similar to news. Inserting metadata into a stream would, in some cases, lower its usefulness. For example, scores and other statistics are bound to a team or player, not a video stream. For this reason, sports statistics are frequently kept in separate databases.

Summary of File and Stream Formats

Basic Stream Formats

Compressed stream formats such as DV and MPEG program streams are well suited for the transport and storage of uncut shots. MPEG program streams are the most popular file format due to the success of the DVD, which is based on this format. DV is used for camcorders and in news applications.

National networks have been built using nothing more complex than a basic stream format. In this architecture, all metadata must be managed in one or more databases; a viable approach to solving real-world problems.

SMPTE 268M—Digital Moving Picture Exchange (DPX)

This format was developed several years ago to support the transfer of uncompressed images between telecine machines and later used for synthetic image file transfers. DPX was standardized in 1994 as SMPTE 268M.

The images that make up a SMPTE 268M file are uncompressed video. SMPTE 268M also supports metadata needed to specify the details of the image size, image shape, sample (pixel) information, and some fixed metadata. A SMPTE 268M file includes space for user-defined metadata.

SMPTE 268M will continue to be used for uncompressed image transfers. This standard does not support compressed video types and other features that are incorporated in newer formats. Although the cost of supporting uncompressed images in a new format is low, it is not clear that a new format would offer sufficient advantages (value) to overcome the expense of converting or retiring existing systems.

SMPTE 360M—General Exchange Format (GXF)

Many broadcast facilities currently use SMPTE 360M. Time code, audio, and compressed video material is sent as a time-multiplexed stream. A simple edit decision list, user data (metadata), and a frame lookup table precede the content. Any material that is currently not in use (handles) is sent after the active material is transferred.

This format is based on packets containing either a frame of video, about a half second of audio, or several seconds of time codes. Each packet has a header that specifies the packet type, as well as identifying the track, the media type, and other information.

The "user data" specified by SMPTE 360M is metadata. It can be bound to the entire composition, a specific track, or a segment of a track. The metadata is encoded in a manner that allows the encapsulation of KLV, XML, or other metadata formats.

The frame lookup table was designed to optimize partial file retrieval operations. This is used when a SMPTE 360M file is stored on data tape or on a disk-based archive. Server-to-server partial file transfers are also supported.

The original version of SMPTE 360M documents uncompressed audio, motion JPEG, DVCPRO, MPEG video, and multiple tracks of each type can be part of the stream. A revision of this standard has been submitted to SMPTE. The revision includes enhancements such as high-definition media types and frame rates, AC-3 audio, Dolby-E audio, and others.

Advanced Authoring Format (AAF)

AAF supports a wide range of features needed for post-production and high-end editing applications. The AAF Association has released an application programmer interface (API) specification and a Software Development Kit (SDK) that includes a reference implementation. The SDK was derived from the OMF initiative. Several vendors have adopted AAF.

The reference implementation can be used by product developers to ensure better interoperability between different vendors' products. The reference implementation is not a substitute for written specifications; AAF offers both.

The post-production process frequently uses many different devices. Video editing is performed on one system, special effects may be rendered on a different device, and audio work uses different systems. AAF files can be interchanged between all these devices while retaining a complex EDL and rich metadata. AAF's ability to save and interchange post-production projects as work-in-progress is key.

Material Exchange Format (MXF)

The proposed MXF standard provides features for many broadcast applications including acquisition, contribution, operations, archives, and D-Cinema. In one MXF application, files will be transferred into and out of AAF-based editing systems. File transfers between AAF devices will probably use the native AAF format.

The MXF specification includes "operational patterns" that describe several layers of features. The simplest pattern is an uncut stream with higher levels adding editing features. The current design includes metadata-only files.

The Pro-MPEG Forum started designing MXF and submitted the work to SMPTE for standardization. MXF is making reasonable progress for such a complex standard. The SMPTE standards process encourages comments and positive contributions. Incorporating feedback may slow completion, but results in a more useful and complete standard.

Several vendors and users are participating in the MXF development effort. An industry-wide collaboration does not always move quickly; however, MXF has the potential to become a widely accepted format.

Internet Streaming and PC Formats

Audio and video are important applications on today's personal computers. Microsoft Corp., Apple Computer, RealNetworks, and others have developed formats and tool sets to support audio and video on PCs.

The feature sets in these formats traditionally focus on desktop systems, not broadcasting. Since the personal computer market is large, a significant number of applications are available. Some desktop applications meet broadcasters' needs today and more may in the future.

One Size Will Not Fit All

The number of features a given stream or file format supports varies. As designers try to solve more problems with a single format, the complexity increases dramatically. Designers sometimes invent simple solutions to broad problems, but unfortunately, this is not always the case.

Formats with low latency that support simple edits such as video cuts and audio fades in/out, are required for applications such as news, sports, and on-air operations; they are not adequate for post-production. For an archive, simplicity and full disclosure of the format are critical. Partial file retrieval is also important.

Conclusion

Several file and stream formats are currently in use, with more being developed and standardized. Broadcasters must select formats that meet their needs.

Basic stream formats are optimized for acquisition, contribution, and emission: AAF is designed for post-production and rich editing applications; MXF is being developed to meet the needs of several applications; SMPTE 360M is used for news, sports, on-air operations, and archives; and SMPTE 268M enjoys great success for uncompressed image file transfers.

Today, broadcasters must match file and streaming formats with their workflow requirements and available products. In the future, many servers, editing systems, and other devices will support multiple formats, thus simplifying the integration of large broadcast systems.

Bibliography

David J. Bancroft, "Recent Advances in the Transfer and Manipulation of Film Images in the Data and HDTV Domains," *SMPTE J.*, *109:*273-286, Apr. 2000.

David J. Bancroft, "Images and Formats," *SMPTE J., 110:*579-585, Sept. 2001.

SMPTE 268M-1994—File Format for Digital Moving-Picture Exchange (DPX), *SMPTE J., 102*:760, Aug. 1993.

Ray Baldock, "Enabling Network Interoperability Between Video File Servers, The SMPTE 360M General Exchange Format (GXF)," *Proc. 143rd SMPTE Tech. Conf.*, New York, NY, Nov. 2001.

SMPTE 360M-2001—General Exchange Format (GXF), *SMPTE J.,111*:315-326, May 2001.

Brad Gilmer, "Enabling Better Media Workflows—An overview of the Advanced Authoring Format," www.AAFassociation.com.

James H. Wilkinson, "An Overview of the Material Exchange Format (MXF) Specification," *Proc. 143rd SMPTE Tech. Conf.,* New York, NY, Nov. 2001.

Acknowledgments

The author would like to thank Ray Baldock and Stephen Burnside for their contributions to this work.

THE AUTHOR

Bob Edge is a principal engineer for Thomson Broadcast Solutions/Grass Valley. He is currently responsible for enhancements on the Emmy award-winning Profile line. During his career, he has worked as an engineer, architect, project leader, and program manager on several products.

Recently, Edge led the effort to standardize the General Exchange Format (GXF), now known as SMPTE 360M. He is active in television standards work, including participation in SMPTE, a few European Broadcasters Union projects, and the Pro-MPEG Forum. Previously he participated in the EBU/SMPTE Joint Task Force.

Edge has a bachelor's degree in computer science from Oregon State University and is the author of several technical papers.