FILE INTERCHANGE HANDBOOK

for images, audio and metadata

Editor in Chief BRAD GILMER

DPX, GXF, MXF, AAF, QuickTime, Windows Media 9–ASF

Focal Press

File Interchange Handbook
For Images, Audio, and Metadata

This Page Intentionally Left Blank

File Interchange Handbook

For Images, Audio, and Metadata

Editor in Chief,

Brad Gilmer

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Focal Press is an imprint of Elsevier

Focal Press

Focal Press is an imprint of Elsevier
200 Wheeler Road, Burlington, MA 01803, USA
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

∞ Recognizing the importance of preserving what has been written, Elsevier prints its books on acid-free paper whenever possible.

Library of Congress Cataloging-in-Publication Data
Application submitted.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN: 0-240-80605-0

For information on all Focal Press publications
visit our website at www.focalpress.com

04 05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Contents

Introduction

It was not long ago that the average computer had a 200 MB hard drive, a 50 MHz processor, and a connection to the outside world that consisted of either a 3.5" floppy disk or a 9,600 band modem. If the computer was connected to a network, which was rare, the network consisted of a simple hub operating at 10 Mbps. A single, large file transfer from one computer to another brought the entire network to its knees. In this environment, it seemed ridiculous to move a project from a video tape to a computer, where it could take at least 200 GB of storage. It was equally ridiculous to move film images to such a computer environment. There were a few digital film facilities, but they employed special computers and huge storage systems. To sum up, just a few years ago, computers and networks could not support professional imaging applications.

Fast forward to 2003: You cannot purchase a hard disk smaller than 80 GB. Processors run at 1 GHz or faster. Floppies are almost obsolete, having been replaced by keychain USB drives and gigabyte PCMCIA cards. Many businesses have at least a 1 Mb DSL connection, and networks are deployed at 100BaseT or gigabit speeds for a few hundred dollars. Clearly, this is a different environment from just a few years ago. But one additional advance has been crucial in enabling imaging applications on the desktop—compression. Using current compression technologies, that 200 GB file is now 8 GB. New compression technologies promise to reduce file sizes even more. Today, it does not seem unrealistic to transport images, sound, and metadata on a computer network. The promise of doing so on a generic computer platform is quickly becoming a reality. Given these advances, the time is ripe for the development of file formats for our industry.

But why not use generic information technology (IT) file formats? Why continue the expensive and, some might argue, arrogant past by creating additional industry-specific file formats? The industry is using many computer-related technologies, and that trend is increasing. However, some requirements are not met by existing generic formats. Take partial file transfer, for example. This feature comes in handy when you want to retrieve two frames of video from an 8 GB file. Another requirement is metadata support. Many of the file formats in this book have metadata support tailored for imaging applications.

In addition, many of the formats in this book represent a move by the industry from purpose-built television infrastructure to more commodity-driven IT infrastructure. To the extent that the needs of the user can be met by a commodity IT product, it is good news. Users and manufacturers alike can take advantage of lower-cost commodity computer products applicable in our industry. Although we still may require some specialized file formats, we can use these file formats on commodity technology platforms.

This book covers all major file formats used for the professional interchange of images, sound, and metadata. Typically, these file formats are used to transfer content from one system to another in a film, postproduction, or broadcast facility. Out of the thousands of available file formats, the ones in this book were selected either because they are established in the industry, or because they are in development and have features targeted toward professional applications.

The industry has made a huge financial commitment to the development of these file formats. It has spent millions of dollars and thousands of hours developing the formats in this book. Clearly, the industry believes these formats are extremely important.

This book is an interesting study in the advancement of technology and ideas. Digital Picture Exchange (DPX) was invented before the concept of metadata became popular. It envisioned a world in which the transfer of metadata with the image would be critical. The work of the EBU/SMPTE Task Force was seminal in establishing a roadmap for the future. In retrospect, it was successful despite its long name (The EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Programme Material as Bitstreams). The task force caused the reorganization of the Society of Motion Picture and Television Engineers (SMPTE), and significant work began on the standardization of several technologies, which led to many of the formats covered in this book. The General Exchange Format (GXF) built upon previous experience in the industry and took into account many advanced user requirements, such as partial file transfer and index table support. QuickTime and the Advanced Systems Format (ASF) come from the computer industry and have built on lessons learned there. The Material Exchange Format (MXF) spans the divide between the tape and the computer worlds with support for streamable content and enhanced metadata. The Advanced Authoring Format (AAF) draws from both camps, combining computer technology with extensive metadata support for film and postproduction workflows.

This book collects the writings of many of the best authorities on the subject of professional file interchange. In most cases, the authors contributed significantly to the file formats themselves. Many of the authors have a long history in the motion picture and television industry, but not surprisingly, some of them come from the computer industry, which has exerted a growing influence on the world at large and has changed the world of professional production.

This book does not cover nonprofessional file formats. That does not mean such formats do not have a place in the professional environment. It just means that the Editor elected to keep the scope of the book at a manageable size by limiting the subject to professional formats. For the most part, this book also does not cover streaming formats. Other books cover streaming in great detail. This book is focused on the movement of images, sound, and metadata via file transfer from one system to another. Finally, this book does not cover the application of file formats. This may be added in a future edition.

Much of the work represented in this book would not have been possible if some enlightened people in the industry had not tackled issues surrounding the convergence of IT and content production. I would like to thank all the participants of the European Broadcast Union (EBU)/SMPTE Task Force report for their landmark efforts in this area. I would also like to thank Merrill Weiss, who had the original concept for this book. Merrill has made many contributions to the industry over his career. In addition, Mike Cox has been involved in file formats for several years. Mike was one of the primary reviewers of this book, and I thank him for his time and effort.

This Page Intentionally Left Blank

This book is dedicated to its contributors. They are the best in the industry, and it is an honor to work with them.

This Page Intentionally Left Blank

1 CHAPTER

Convergence of Information Technology and Traditional Television Production

Hans Hoffmann

1.1 INTRODUCTION

The process of television program production has changed dramatically because of several technological and economical alterations. The often-cited convergence between Information Technology (IT) and traditional television production not only enabled television facility designers to benefit from increased processing power and storage capacity but also created the need to address radical changes in workflow. At the same time, challenging economic concerns have pressured broadcasters to increase efficiency, reduce production costs, and establish new businesses.

Information technology offers ways to achieve these goals. It enables significantly enhanced creativity, improved efficiency, and more reasonable economies of scale in the creation of television programming. Driven by the growing demand for programming to fill the multiplicity of competitive distribution channels to consumers being installed around the world, IT likely will become a pervasive force in teleproduction.

The term *IT-based production* (television, audio/sound, and motion pictures) has become common in the industry. The following are some of its major characteristics:

- Content is handled in file form and can be transferred in non real-time.
- Content is transported via standard IT networks and protocols.
- IT-based production relies on the creation and use of metadata as computerized information.

- IT-based production relies on the use of identifiers (i.e., Unique Material Identifiers, or UMIDs).
- IT-based production requires significant efforts in information management (content and asset management systems).
- IT-based production comprises an end-to-end solution (e.g., from acquisition through editing to playout).

The digital transformation of the production chain, which started in the 1990s, was characterized by two significant technical standards: International Telecommunication Union-Radiocommunication (ITU-R) Recommendation BT.601 (studio encoding parameters of digital television for the standard 4:3 and wide-screen 16:9 aspect ratios) and the Serial Digital Interface (SDI) according to ITU-R Recommendation BT.656 (interfaces for digital component signals in 525-line and 625-line television systems operating at the 4:2:2 level of recommendation under ITU-R BT.601, Part A). These baseband standards were developed by user organizations and the industry in a worldwide initiative. By applying a sampling rate of 13.5 MHz for the luminance channel (Y) and 6.75 MHz for each of the color difference channels (B-Y and R-Y), the resulting bit rate of 27 Megawords per second (10-bit resolution) necessitated a serial bit rate of 270 Mbps. Traditional digital television studio installations apply these standards for the digital encoding and infrastructure (backbone) to exchange TV signals. After the ratification of Recommendation 601 and 656, an entire industry (the traditional broadcast vendors) became established in the market and developed successful products. Several manufacturers provided components such as signal routers for moving Recommendation 601 and 656 signals between the devices in the studio environment. Others developed videotape recorders such as the D1 (8-bit uncompressed) and the D5 (8- or 10-bit uncompressed). At that time, the IT industry played a less significant role in the professional area of production. The reason was that the bit rate of 270 Mbps made it difficult to provide cost-effective products, in particular for moving 270 Mbps signals over IT networks and for affordable hard-disk storage (e.g., the storage requirements for one minute of Recommendation 601 in 8 bit would be about 1.6 GB). However, users and the industry soon discovered that, by applying digital compression to the baseband signal according to Recommendation 601, both the large storage requirements for uncompressed ITU-R BT.601 signals and the transfer bit rate could be reduced. An obvious drawback was that different, incompatible compression algorithms were introduced into the professional production environment (unlike MPEG-2 compression, widely adopted for distribution to the viewer). Most of the first server systems in the 1990s used the Motion JPEG compression algorithm. For videotape recorders, prominent examples were the Digital Betacam (compression based on Discrete Cosine Transform, or DCT) and, more recently, the IMX (compression based on MPEG-2) or the DVCPRO (compression based on digital video, or DV). Most compression algorithms (e.g., DCT, DV, and MPEG) apply methods to reduce the bit rate by deleting information from the original signal. This information usually cannot be recognized as missing, because of nuances in the human visual system. (Irrelevant information reduction also is called lossy coding.)

The existence of incompatible, and sometimes proprietary, compression algorithms implemented in products from competing vendors resulted in discord. For example, videotapes (apart from different cassette size and recording technology) could not be exchanged across vendors because of different compression systems. The solution to this problem was to decode from the compressed domain to ITU-R BT.601(baseband, uncompressed) to use the Recommendation 656 interface to transfer program material (e.g., video) between equipment from different manufacturers. This decoding and reencoding resulted in a loss of picture quality. Considering the whole production chain with all of its systems, multiple decoding and reencoding (or generations) led to significant impairment of the picture quality. The first step in avoiding such picture-quality loss was the development, standardization, and finally, the market introduction of the Serial Data Transport Interface (SDTI, standardized in SMPTE 305M). This interface was developed to transport packetized data such as compressed video within a studio production environment in its native form. The transport mechanism is compatible with SDI (ITU-R BT.656 or SMPTE 259M). This helped to avoid picture-quality reduction during compressed content exchange—as long as the source and destination systems supported the same compression format.

In addition to the advantages that compression technology has delivered to the traditional broadcast environment, nonbroadcast industries—computer, IT, or multimedia, in particular—provided solutions for the professional TV production market. This significant development typically is referred to as the *convergence* between traditional broadcast and IT (computer) industries. In addition, with rapid developments in storage capacity and signal processing power, computer equipment became a better candidate for replacing traditional video-production equipment. This has led already to dramatic changes in workflow and the way programs are produced. Nonlinear editing (NLE) became affordable with computer/server-based storage systems and has already started to penetrate almost every broadcaster. Compression, as applied today in many professional TV products, has enabled not only efficient storage on different media but also the transfer of content via computer networks in non real time, in real time, or faster than real time. Figure 1.1 gives an overview of the storage and interface technologies in professional TV production. It also shows the effect of compression on interface technologies—first on the development and

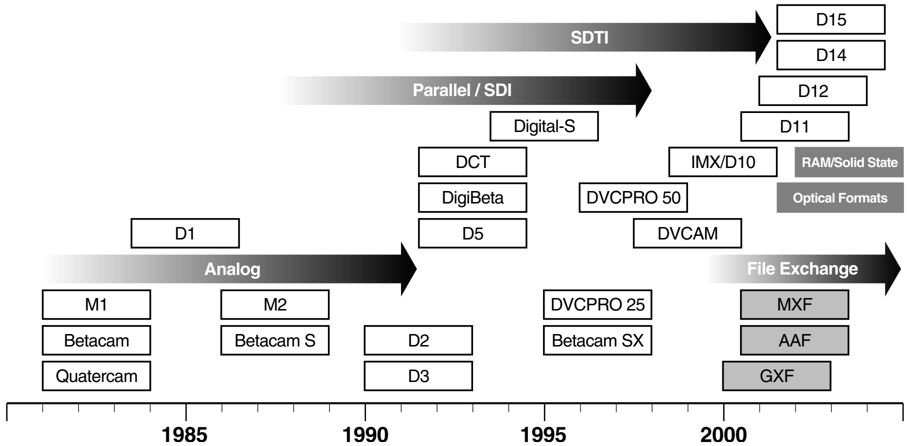

FIGURE 1.1 Recording and interface technologies in television production

standardization of SDTI, which enables the exchange of compressed video signals via SDI-based studio infrastructure, and later by introducing file exchange via computer networks or removable media (e.g., optical disc).

1.2 HISTORY OF FILE FORMATS IN PROFESSIONAL CONTENT PRODUCTION

As outlined, computer-based technology has increasingly proved its usefulness in many applications within the professional broadcasting environment. Examples can be found worldwide in server systems for production, postproduction, playout, and archiving. Users and vendors are witnessing the first attempts to extend the application range of hard-disk, optical-disc, and memory-based storage to mobile applications in professional news gathering. In accordance with this, broadband networking, including continuous recording of transmitted television programs at home, has started to penetrate the consumer domain.

The common denominator in all these applications is the transport of program content and its storage on nonlinear media or via networks within proprietary file formats. Therefore, program exchange can only be carried out across platforms that can manage and exploit such proprietary file structures. Users have already expressed a strong requirement to share files between systems made by different vendors. Sharing in this context refers to the exchange of content assembled in files by means of removable media or, in particular, by directly accessing the content stored in these files through standardized interfaces and network protocols. The operational and economic benefits of sharing files using nonproprietary file formats are summarized as follows:

- Multiple users can simultaneously access data related to a common project within a distributed production environment.
- File exchange does not degrade picture quality because the compressed video in the file body can be transferred in its native compressed form.
- File exchange can be carried out through local and wide area networks (LANs and WANs) at different speeds (i.e., slower than, equal to, or faster than real time).
- The speed of the file transfer can be adapted to the available channel bandwidth. (If the network allows 10 Mbps, the file is transferred at that speed; if a faster network is available and the peripheral equipment can support it, then the file can be transferred at a higher speed.)
- Users can balance the transfer costs against the transfer time.
- Metadata, audio, video, and data can be transferred in one wrapper.
- The physical media (tape, optical disc, etc.) can be separated from the content embedded in the file.
- A horizontal system, following a layered model, is possible.
- Broadcast systems can be built using readily available computer equipment that might result in lower overall system cost.

Within a distributed, multiuser environment, these advantages can only be exploited if the source and destination system can interoperate. This requires the file format and its content to be well defined and open. However, most implementations available to the broadcast market have employed different proprietary file formats. Some of them have been directly adopted from the IT industry, such as QuickTime, Audio Video Interleaved (AVI), and Advanced Systems Format (ASF), whereas others have been developed for more demanding applications in the professional broadcast world, such as MPEG and GXF. These professional applications have been successfully standardized.

Unfortunately, the professional video market has faced challenges in adopting IT standards. Specifically, there have been challenges with the numerous, incompatible, or nonstandardized file formats of the IT world. In addition, most IT file formats had difficulties complying with the emerging needs of the professional broadcast industry (e.g., file size, editing capabilities within a file, and payload neutrality). Nevertheless, the enthusiastic introduction of server-based NLE stations in most broadcast installations worldwide and the requirement to interconnect different NLE stations has created a need for common, standardized, and open file formats that can cope with all the requirements of professional production. The major criterion is the availability of file formats that permit the exchange of information in its native form—such as compressed or uncompressed video, audio, data, or metadata (content). This exchange occurs between different systems as files, independent of the location of the users (i.e., distributed production). Initially, many supposed that a single, standardized file format would meet the needs of the entire postproduction and broadcast community. This view had to be quickly corrected. First, professional media production involves features that vary significantly between application environments. Second, depending on the viewpoint of the user or manufacturer, a single file format standard would be either over- or underdesigned for their requirements. Third, the preexistence of files in large installations and successful standardization of these existing file formats cannot represent state-of-the-art and future-proof demands (because it may already be obsolete technology).

Taking into account this situation, several accredited organizations in the professional TV production world—such as the European Broadcasting Union (EBU), the Society of Motion Picture and Television Engineers (SMPTE), and the Association of Radio Industries and Businesses (ARIB)—as well as the Pro-MPEG Forum and the Advanced Authoring Format (AAF) Association, have started to develop a new file format called the Material Exchange Format (MXF). This format addresses the user requirements for mainstream, IT-based, TV program production such as news, archives, and production. Substantial contribution came from users in the United States, Asia, and especially Europe. The development of MXF also had strong and broad industry support (via the Pro-MPEG Forum and projects of the European Commission). It has established a degree of interoperability with AAF for the exchange of information between the graphics and postproduction environments.

1.3 MIGRATION TO SERVER- AND NETWORK-BASED TECHNOLOGIES

Users will be faced with several technology choices when introducing IT-based system components, such as computer servers or networks. Choices include signal and coding formats for acquisition, contribution, archiving and production of the methods of interconnecting systems; the management and control systems of these systems; and the file format used for storage and transfer. Usually, any computer or server-based system that stores television signals on nonlinear media (hard disk, optical disc, etc.) or even on data tape handles television signals in file form. An NLE system, for example, would convert incoming SDI or SDTI streams immediately to a file for storage or other application purposes. Once files have been created, it is natural to want to maintain and exchange these files via networks or on removable media. Since convergence began, several interface technologies have been offered for content exchange. Real-time, uncompressed digital video signals in the studio for standard digital television are still best handled by ITU-R BT.601 and ITU-R BT.656 SDI. In some applications involving the transfer of compressed video, users may wish to use the SDTI derivate of SDI, which offers not only the exchange of compressed signals (without picture-quality loss) but also faster-than-real-time transfer (e.g., four times 25 Mbps). Similar traditional interfaces are available and used for high-definition television (e.g., SMPTE 292M for uncompressed HDTV). The IT industry has been offering several interfaces or networks originating mainly from the telecommunication or multimedia market (Fig. 1.2). Here are some examples:

- IEEE 1394: This consumer point-to-point interface supports up to 64 nodes at speeds up to 800 Mbps. Its main application is in the real-time (and faster-than-real-time) transfer of compressed video signals. The SMPTE also has standardized the mapping of the DV-based compression scheme. In the professional domain, its application can be found in the desktop environment of VTR-to-VTR, VTR-to-computer, and hard-disk interconnects.
- Fibre Channel: This LAN-oriented network supports data rates up to 10 Gbps, mainly for the high-speed interchange of files. It also supports large storage arrays and server systems. Fibre Channel has become a state-of-the-art technology for internal and external interconnection of professional studio servers.
- Gigabit Ethernet: This uses chip technology similar to Fibre Channel and is increasingly a cost-effective LAN technology for medium speed applications in professional studios. It seems to be an attractive technology for users who want to migrate from low-bit-rate Ethernet to higher speed networks.
- ATM: This technology has become one of the dominant technologies for wide area interconnects. It supports high speeds (e.g., 622 Mbps, STM-4), and offers different transfer modes (quality of service, or QoS) for real-time and file-transfer applications.

Internet Protocol (IP), with different protocols such as Transmission Control Protocol (TCP) or User Datagram Protocol (UDP), runs over some of the interface technologies just described. It can also be thought of as a transport mechanism. Traditionally used for on-demand, non real time file transfer, the

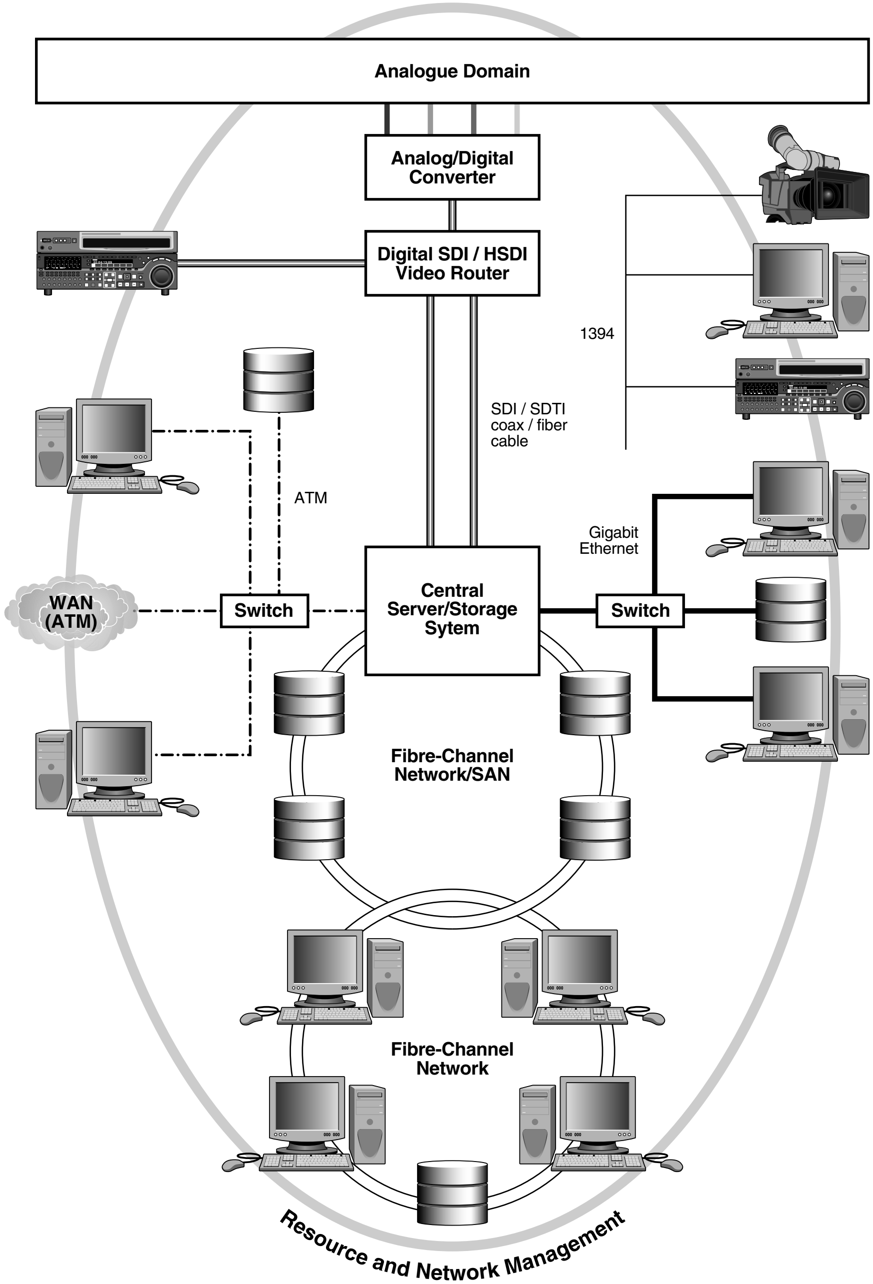

FIGURE 1.2 Network technology scenarios in TV studios

multimedia revolution has added streaming over IP. *Video over IP* is a term that has been used in this context for many applications, such as for low-bit-rate MPEG transport in videoconferencing and content browsing. The major uses of IP in the professional broadcasting domain, however, have been the exchange of files and search and retrieval in newsroom- or archive-browsing applications. At the beginning of its use in the professional media environment, some proprietary modifications were introduced to TCP/IP (tuned TCP buffers) and its application program interface (API), known as the File Transfer Protocol (FTP). This allowed compliance with the challenging requirements of gigabyte file exchange and avoided congestion or packet loss. Today, the Internet Engineering Task Force (IETF) provides additional specifications (Request for Comments, or RFCs) to meet the demands for handling and transmitting large files (e.g., via "Jumbograms"). Challenges remain, including the issue of signal loss during dynamic rerouting that may happen across public networks. These challenges are being actively addressed.

Most interface technologies proposed by the IT world for professional program production were designed to operate in asynchronous modes with files, rather than with high-bit-rate, synchronous, real-time signals. This is no surprise considering that almost all original Internet traffic came from on-demand (non real time), file-based business and military applications. As a consequence, the broadcast industry is first addressing file-based applications as it moves to adopt IT. Eventually, many of today's applications, such as time-consuming VTR dubbing, may be accomplished using file transfer (preferably, faster than real time). Centralized storage of content, either using Storage Area Networks (SAN) or Networks Attached Storage (NAS) applications, will greatly reduce the need for file copying—if distributed file systems are applied.

Broadcasters need real-time transmission of content, and the IT world is offering solutions. The IETF, for example, provides additional specifications to facilitate streaming video over IP via protocols such as the Real-Time Transport Protocol (RTP), the Real-Time Control Protocol (RTCP, specified in RFC 1889, 1890, etc.), and the Resource Reservation Protocol (RSVP, specified in RFC 2205). There is even an RFC available that deals with the transfer of serial high-definition signals (SMPTE 292M) via IP.

1.4 GOING DIGITAL, BIT STREAMS, AND FILES: THE EBU/SMPTE TASK FORCE

In the mid-1990s, the two largest forums on professional broadcasting, the EBU and the SMPTE, began to study the effect of IT on professional television program production. In mid-1996, EBU findings[1] stated that the performance, stability, and reliability of traditional television production can only be met using IT technology if users insist on open and standardized technologies. The SMPTE, a standards-setting body for professional broadcasting and production, in an independent study came to similar conclusions. The organizations have worked closely together over the years. During the International Broadcast Convention (IBC-Amsterdam) on September 12, 1996, the two groups decided to establish the EBU/SMPTE Task Force for Harmonized Standards for the Exchange of Program Material as Bit Streams (later referred to as the Task Force). The Task Force was charged with two assignments: "a) to produce a blueprint for the implementation of the new technologies, looking forward a decade or more, and b) to make a series of fundamental decisions that will lead to standards which will support the vision of future systems embodied in the blueprint." Two significant reports were published as a result of the joint effort.[2]

In carrying out its work, the Task Force divided its effort into five fundamental areas; compression, physical link and transport layers for networks, wrappers and file formats, metadata, and file transfer protocols. Each area was assigned to a dedicated subgroup. Major aims of the Task Force were to provide a framework, principal architectures, and a structured point of view for future IT-based television production (layered approach); to identify suitable technologies; to provide stability via standards; and to initiate new developments.

1.4.1 Interoperability: the Value and Need for Standards

Some users have been concerned that the migration toward IT in digital video systems could result in the abandonment of specific industry standards, such as those of the International Telecommunication Union (ITU), the SMPTE, and the EBU. For that reason, users tried to establish joint efforts to increase their effect on the market and to initiate standardization in several areas. It may be useful to consider standardization efforts as providing a well-balanced "force" to vendors. Attempts to over-standardize a technology might be cost-intensive and might hamper competitive products as they enter the market. In the case of well-accepted and proper standardization, users and the industry may need to specify additions on top of a standard to meet the last 10% of functionality that broadly based commercial systems were never designed to meet. It is important for the market that a variety of systems can be set up by "mixing and matching" products from different vendors. An example of the need for clearly defined best practices is the use of MPEG-2 4:2:2P@ML. MPEG provided the baseline standards, but the SMPTE provided additional recommendations and standards (e.g., SMPTE 356M) to define the bit-stream usage in professional production environments. By carefully selecting "nominal" values from the ranges of choices within a standard, users can better achieve interoperability for their individual and sometimes competing applications. In other words, it may be that the role of professional broadcast standards organizations is a combination of writing standards when none exist and advising the specific broadcasting industry segment about applicable IT standards and how they can be used in professional video applications.

As a general design philosophy, user organizations should attempt to choose open standards to which all qualified vendors can design and supply systems. Under such conditions, vendors would be able to develop products that can interoperate, but they could differentiate themselves in the application functionality. (Users would have the benefit of selecting products from a wider range of manufacturers.)

The EBU has published a document titled, "Statement on Open Standards (D79)" that emphasizes this requirement.[3] Similarly, the SMPTE defines, in its "Recommended Practice RP 213 Annex B," the meaning and levels of interoperability.[4] This definition of interoperability levels became important in guiding the development of standards, because of increasing complexity of systems and the adoption of horizontal system designs. For example, a single standard may define interoperability for a particular layer, such as a compressed bit-stream syntax, but this standard does not guarantee that the file within which the compressed bitstream is wrapped can be opened by the target application. Consequently, additional standards are required (in this example, the file format and the standard to map the bit stream into the file format) to achieve interoperability of a certain application.

By selecting international standards wherever possible, global competition can be maintained, providing all international players with opportunities to contribute their technologies to common systems and data exchange. Often, the nomenclature of compatibility, or interoperability, and standardization is used in a similar way. Note, however, one clear distinction important for the niche market of professional TV production: Products that are interoperable or that can interchange content in a compatible way, such as via a common file format, may increase their value if this interoperability is achieved using a standard developed from an accredited and ratified standardization body. This will ultimately assist long-term stability.[5] For example, files with compressed content may be stored in archives and might be accessible by today's products. Nevertheless, only a well-defined standard, describing the technical details of the file and how to decode the compressed signals, will allow users to access the material over time. Applying these considerations to the Task Force and the subsequent work on file formats in the SMPTE, the EBU, and other bodies (such as the Pro-MPEG Forum and the AAF Association), the ideal result of a standardized file format

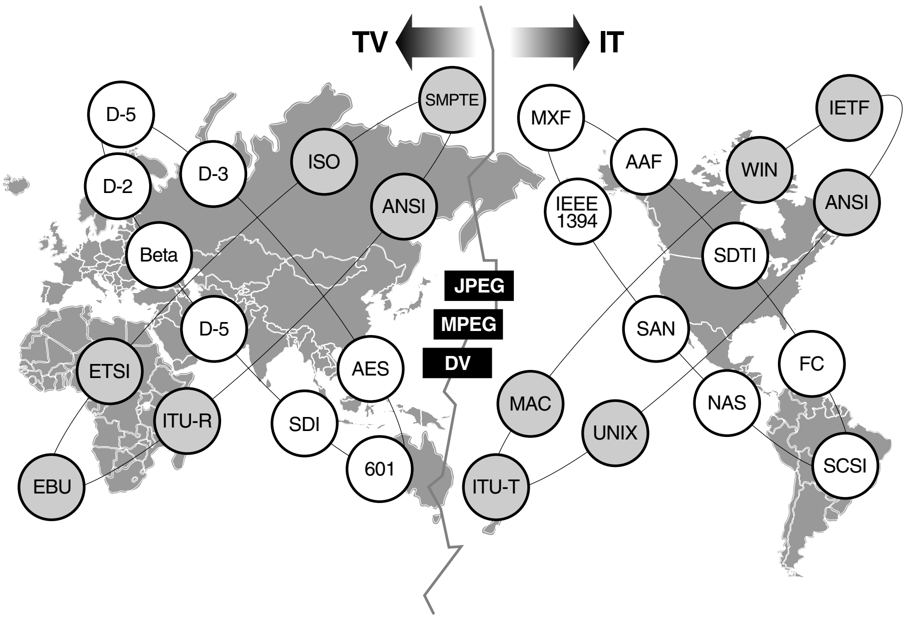

FIGURE 1.3 Worldwide standards bodies

would be a TV program encapsulated in a file format that can be exchanged between different systems.

The professionals in the broadcast world had to painfully learn to deal with standards-setting organizations originally set up to serve different markets—the telecommunications or IT world. In this context, the major challenge is to achieve mutual understanding about the requirements for technology, workflow, and processes when developing or adopting standards (Fig. 1.3).

1.4.2 Problem Description and Layered Approach

Television production systems of the future will integrate news environments, production, postproduction, graphics, and archives more tightly than current installations. In particular, archives will need to be open to different forms of content, and they will have to embrace both the international broadcast community and the multimedia industry. Moreover, metadata or information about the content will be as important as the video and audio itself. The Task Force came up with the following formula:

$$CONTENT = ESSENCE + METADATA$$

Here, essence represents video, audio, and data; metadata represents information about the essence.

After the introduction of content and asset management, this equation was broadened as follows:

$$ASSET = CONTENT + RIGHTS$$

The metadata part is associated with rights information. The logical statement here is that, as a broadcaster, if you have content but you do not have the right to use it, then it has no value—it is not an asset. In fact, it may be a liability.

In future IT-based TV production scenarios, large content-information storage systems will become central to the production process and will need to be managed in an efficient way. They will become the core elements of news and production environments and will likely be the central storage and management system for the entire production process. This is a major shift in view on the role of archive systems. Traditionally, archives have been viewed as an "end of pipe" process. However, with IT-based technology, archives are migrating to a core role in the facility.

In the traditional broadcast production environment, systems were developed with a vertical approach. This made the integration of different information types difficult. Often, solutions from one vendor made it impossible for users to replace parts of their system with products from a different vendor. This "lock-in" to one vendor's products—often associated with proprietary signal or interface technologies—was inconvenient for many users. The IT world, on the other hand, has followed a horizontal or modular approach to systems. This was a natural consequence of being software-centric and dealing with a rapid rate of change. In addition, the rate of change in one area did not keep pace with the rate of change in a different area. This created a strong requirement for the ability to replace individual system components, rather than a whole IT infrastructure. The approach of the IT world has been to follow a horizontally layered model known as the ISO reference model for open system interconnection (OSI-layer model, Fig. 1.4).[6]

In an ideal world, the model permits the exchange of individual layers without affecting the others and, in consequence, provides a framework for the development of standards valid for individual layers. Theoretically, this approach allows a user to upgrade the physical topology of an IT network (e.g., move from Ethernet to Gigabit Ethernet) without modifying all the applications using the network. Applying the OSI approach to file formats in practical broadcast production processes would mean, for example, that a fully standardized file format represents a single, horizontal layer of interoperability.

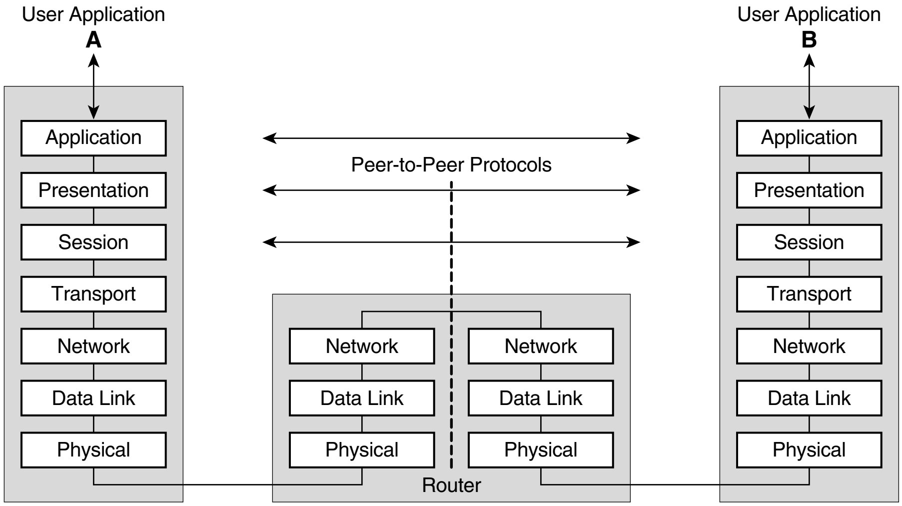

FIGURE 1.4 OSI-layer model

The different compression schemes, different interface and different workflow scenarios have presented problems in all technology layers of professional production systems. The EBU/SMPTE Task Force identified at the outset of its work that the IT-based OSI model would be a suitable methodology for providing solutions for the multilayer interoperability problem in network and server-based (IT) broadcasting environments. On a case-by-case basis, the Task Force used a simplified model of the OSI reference system to organize traditional broadcast technology according to a layered approach. Figure 1.5 provides an example of a simplified OSI model. The model will vary with the broadcast technology used in each of the layers. For example, if an SDI infrastructure is used in the networks layer, the transfer protocol layer would not exist. SDI (a traditional, vertical standard comprising several layers) is a unidirectional interface technology with its own clear framework. For instance, the start and end of active video signaling of SDI could belong to the formatting layer, rather than the transfer layer. If SDTI over SDI is used, it could be argued that this represents either a protocol layer or an addition to the network layer. In a typical IT network example, such as transferring a video file via Ethernet, the OSI-layer model is clearer: A video signal is either compressed or uncompressed and formatted into a file (file formatting). Then, the file is transferred with a QoS (e.g., the FTP application via TCP/IP) via the network.

This example clearly shows the difficulties that the Task Force encountered during its work. Nevertheless, it became clear that interoperability in future professional broadcasting could only be achieved through a well-structured and well-

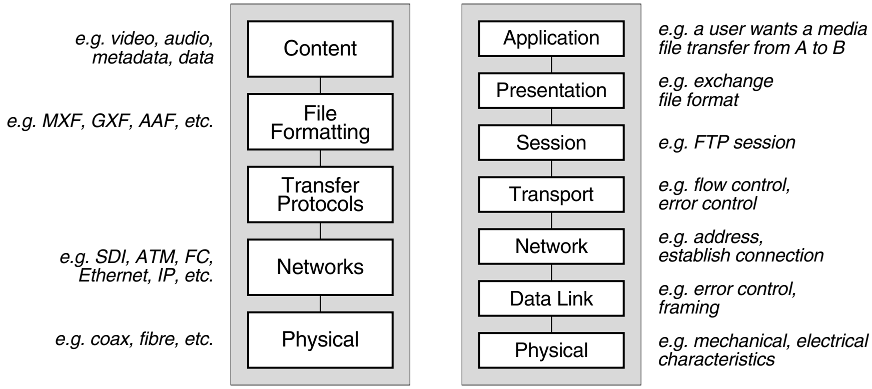

FIGURE 1.5 Simplified OSI-layer models applied to television systems

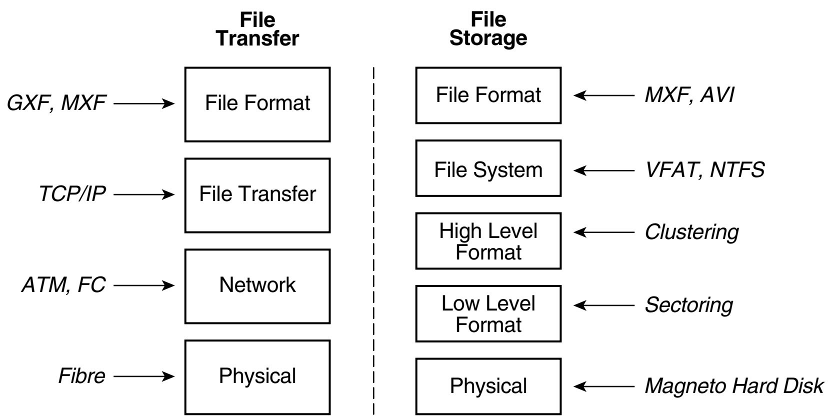

FIGURE 1.6 Storage and file transfer: simplified layer model with technology examples

layered approach, particularly as file formats and file transfer via networks become dominant applications. Another important categorization of the technologies surrounding files and file formats is storage versus transfer of files (Fig. 1.6).

A file is like a container holding all the program elements (content) within a specific project. The file including content can be transferred over a network or stored on a storage medium as a single entity, easily identified by a unique file name for unambiguous retrieval. Once a file is opened, metadata will provide a description of the essence accommodated in the file body and define the relationship (contextual, timing, etc.) of the elements. The definition and standard for a file format is, in principle, independent of the transport mechanism (file transfer) or the storage mechanism for files (layer model, as just explained). However, the often-mentioned requirement to accommodate partial file transfers requires a degree of interrelationship between the file format and the transport mechanism. If a minor transmission error occurs during a transfer of a video file, the network protocols usually initiate a retransmission of the corrupted packets. However, if large parts of the file are corrupted or a user wants to transfer only parts of an existing file (e.g., a few video frames from a large movie file), limited interaction between the file format layer and the file transfer mechanism is required.

A basic file format structure is shown in Figure 1.7. Usually, a file consists of a preamble with run-in sequences, followed by the body and an end-of-file marker. Editorial or descriptive information, such as metadata, typically are located in the preamble. The file body consists of the so-called payload. This can be uncompressed/compressed video, audio, data, or additional metadata.

If a real-time video signal is transferred onto a hard-disk-based server (e.g., for NLE), the incoming data stream is stored as a file. In file transfers between servers over networks, the incoming signal is usually already in file form. It may be directly transferred to the storage medium, or it may need conversion to a different file format before storage. High data throughput, fast nonlinear access to the stored content, and efficient usage of storage are required. Therefore, the file format may need to be restructured to match the inherent file structure and the structure of the segmented format of the storage medium. The latter is called *structured storage*, *low-level storage format*, or *native file format*.

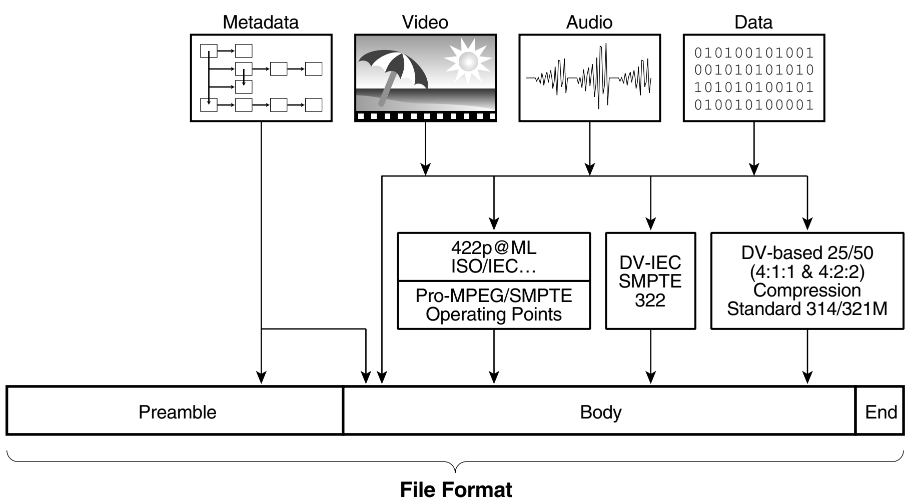

FIGURE 1.7 Example for common signal formats mapped into a generalized file format

The following are the main elements that must be considered when discussing files in professional broadcasting:

- The format used to transfer the information as a file, which may exist only on the wire, may be different from the file format used to store the information on disk or tape
- The storage format that contains the file format used to write the bits to disk or tape
- The file transfer protocols being used
- The API and operating system responsible for generating access to the file stored on disk or tape

Regarding IT-based installations, the discussions on the constraints imposed on moving files between systems have not ended. In particular, the application of streaming files with real-time capabilities generates a challenge for typical IT networks (considering requirements of TV production such as full synchronization, nanoseconds of jitter, etc.).

1.4.3 Results of the EBU/SMPTE Task Force

Systems

To better understand the requirements of system design, the Task Force has developed a model based on orthogonal parameters and intersected by an underlying control and monitoring layer. This model has been used to explore the relationships between signals, workflows, and processes, as well as networks/interfaces, control, and monitoring (management) systems.

The Task Force model can be used to describe or analyze any type of program or activity. The description of part of any system can be made in terms of the model by describing the technologies used to carry each of the planes for any given layer. It can also describe the control and monitoring functions across the activities and planes.

A television system can be considered as several signal-carrying planes controlled by an intersecting control plane. Each production task requires the manipulation of signals in some or all of the planes.

In traditional television systems, the planes consisted of distinct physical systems: Video, audio, and data were carried on different cables. Metadata was often simply written on a piece of paper or tape label. Future systems will not necessarily have these distinct physical systems. Instead, they will be based on

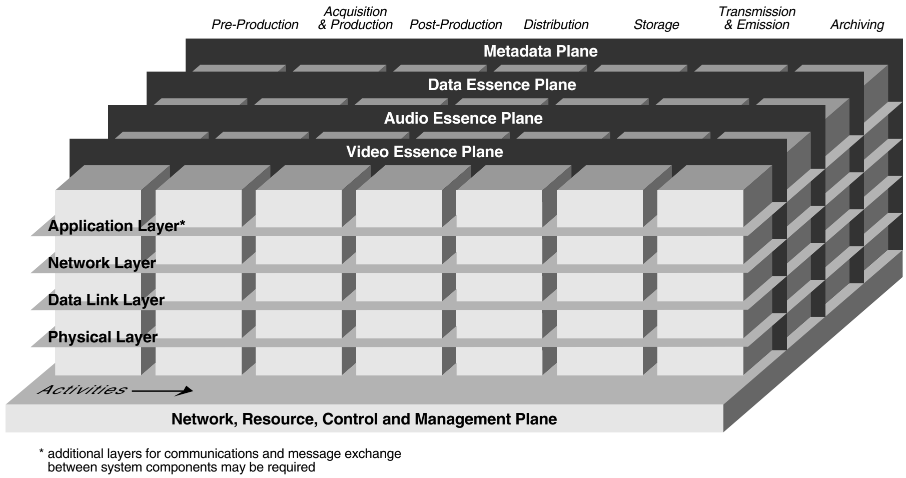

FIGURE 1.8 System model of the EBU/SMPTE Task Force

networks carrying multiplexed signals. It is useful, however, to consider new IT systems in terms of a logical model in which the signal types are distinct.

Figure 1.8 shows the model developed by the Task Force. Recent findings, however, suggest that a further distinction in the application layer is required to address the interaction and message exchange between system components. The additional layers required are Message Exchange Protocols, the definitions of the messages themselves (see, for example, the work of the SMPTE on Management Information Bases, or MIBs), and the API.

Compression

With the introduction of modern compression schemes (e.g. MPEG-2 422 profile and DV/DV-based) in professional production (Fig. 1.9), users have been faced with the following set of key questions:

- Will the data reduction provide the anticipated economic benefits without impairing the picture quality, especially considering the multiple decoding and reencoding required in most production workflows?
- Will the compression algorithm and the bit stream support all operational functions (e.g., editing), and will it be sufficiently standardized so that archived material can be accessed many years in the future?
- Will existing and future interfaces be able to transport the compressed data in an interoperable and standardized way that also allows third-party products to process the compressed signals?

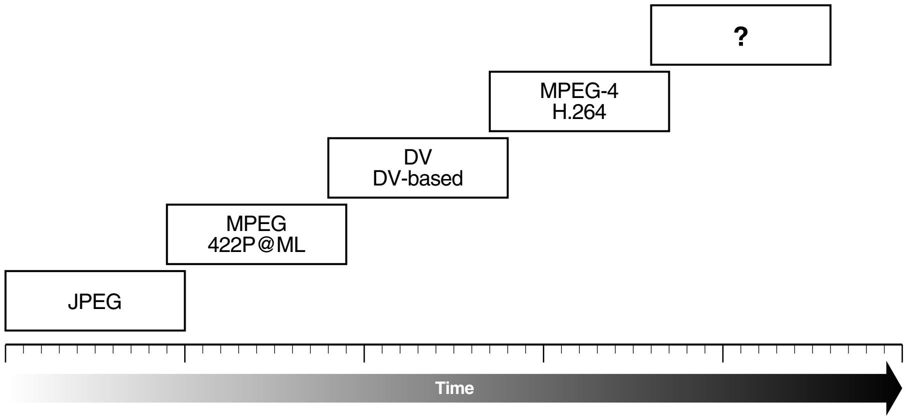

FIGURE 1.9 Evolution of video compression formats

The Task Force findings for audio are as follows: "The EBU/SMPTE Task Force strongly recommends that the AES-3 data stream be utilized for the carriage of all audio signals, compressed or full bit-rate. In some cases, standards will be required to define the mapping of the data into the AES stream." The Task Force further states that the sampling rate will normally be 48 kHz (AES5-1984, reaffirmed 1992), locked to the video frame-rate (AES11-1991) with 16 bits, 20 bits, or 24 bits per sample. With respect to purely file-oriented audio signal processing, the Broadcast Wave Format (BWF) provides an appropriate solution. It is important to recognize that this essential definition for audio needs to be maintained within any file format to be used in professional applications. In other words, a mapping of source signals such as uncompressed or compressed audio, or a mapping of the BWF into any other file format, must ensure that no modification is applied to the source format.

The Task Force findings for video were as follows: For core video applications in mainstream TV production and postproduction for standard definition television, two different compression families on the market were advocated as candidates for use in future networked television production:

- DV/DV-based 25 Mbps with a sampling structure of 4:1:1, and DV-based 50 Mbps (SMPTE 314M) with a sampling structure of 4:2:2, using fixed bit rates and intraframe coding techniques exclusively. DV-based 25 Mbps with a sampling structure of 4:2:0 should be confined to special applications.
- MPEG-2 4:2:2P@ML using both intraframe encoding and Group of Pictures (GoP) structures, and data rates to 50 Mbps. MPEG-2 MP@ML with a sampling structure of 4:2:0 should be confined to special applications.

According to the Task Force, standard definition television in uncompressed form should follow ITU-R BT.601.

In HDTV applications, similar requirements for interoperability have been defined. Currently, dominant compression schemes such as MPEG and DV provide enhancements to cover HDTV applications (MPEG with its high-level profile and DV with its HDCAM and DV-based 100 Mbps derivates).

With respect to the transport of the compressed or uncompressed data in file form, an essential user requirement has been to provide mapping standards. This documentation provides the technical information to map compressed bit streams into the file format in a consistent and interchangeable way. The current standard documents of the SMPTE that represent the mapping of DV, DV-based, and MPEG 4:2:2 compressed video bit streams into SDTI have also been used to define the mappings into the newly defined MXF. (Uncompressed video mapping, according to ITU-R BT.601, is also being developed.)

The functional and operational advantages of adopting a common layer (e.g., the compressed video bit-stream layer) both for mapping into (traditional) streaming interfaces (e.g., SDTI) and for file formats are obvious.

Wrappers and Metadata

The findings of the Task Force in the area of wrappers and metadata provided the foundation for most of the standardization work on professional file formats over the last few years (as will be discussed later in this book) (Fig. 1.10). File formats and wrappers are almost synonyms. According to user requirements,[2] the principal characteristics for wrappers have been defined as follows:

- Wrappers should provide means to convey and link essence and metadata in the logical and physical domain.
- Wrappers must support content organization in the wrapper payload area in a playable form (streaming wrapper) as well as for specific storage or content manipulation purposes (e.g., audio part separated from video).
- Wrappers have to provide links for external data. This can be other wrappers, metadata in a database, essences, and so on.

As a consequence, wrappers (or file formats) have to meet several challenging functional requirements in the different application areas of electronic news gathering (ENG), postproduction, production, archiving, and so on. Further analysis has shown that these user requirements cannot be met by a single wrapper. For that reason, more than one wrapper will be required. An important task for those creating standards in this area was to ensure that future wrappers provide so-called low-processing conversion capabilities. An appropriate exam-