

MXF, un format d'échange pour l'interopérabilité et l'indexation des contenus

MXF est né de deux constats :

- L'arrivée du numérique dans le monde de l'audiovisuel a été une véritable révolution, mais elle n'a pas été le facteur d'uniformisation auquel on a pu penser de prime abord. Aujourd'hui, la diversité des équipements et des formats, la multiplicité des différents processus de traitement du produit audiovisuel tout au long de la chaîne de production, de diffusion puis d'archivage, l'emploi généralisé des serveurs et des réseaux informatiques... ou plus généralement le phénomène de dématérialisation qui caractérise cette évolution ont fait surgir plus nettement encore le besoin de formats communs qui faciliteraient les échanges des documents audiovisuels.

- L'indexation des contenus audiovisuels est une histoire aussi vieille que le monde... de l'audiovisuel. L'étiquetage des boîtes des bandes magnétiques ou des cassettes, la rédaction de fiches cartonnées reprenant les informations les plus pertinentes sur les contenus étaient déjà une forme d'indexation. Mais elle prend aujourd'hui une importance particulière. Des millions de documents vidéo ou sonores sont produits chaque année de par le monde. La dématérialisation des supports, la multiplication des vecteurs de diffusion, la réutilisation totale ou partielle des documents réalisés, les nécessités d'un archivage précis... militent pour le développement d'outils spécialisés qui faciliteraient la gestion des informations sur les contenus.

Face à ces deux constatations, l'idée pour les professionnels fut d'intégrer l'ensemble des composants du média audiovisuel (l'audio, la vidéo mais aussi toutes ces données informatives) dans un unique format de fichier standardisé ce qui permettrait le partage, l'échange et l'archivage des médias, faciliterait la recherche d'information par une description des contenus, assurerait la gestion des droits... et qui serait agnostique vis à vis du mode de compression, du débit ou du format de l'audio et de la vidéo. Deux produits sont nés de ces considérations.

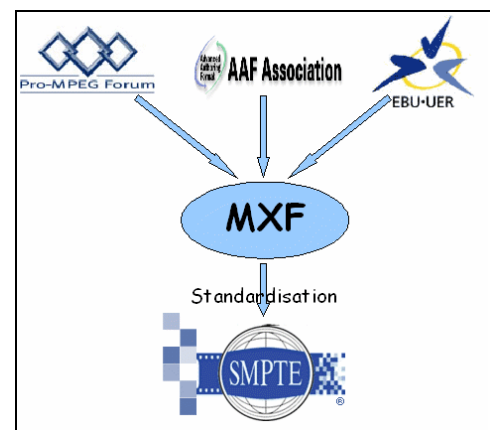
En premier lieu, le format AAF (Advanced Authoring Format), format très (trop ?) ambitieux qui a été conçu pour les échanges en post-production. Il est toujours en cours de développement et non encore normalisé. Nous ne l'évoquerons pas ici.

Faisant l'objet de travaux plus récents, MXF (Material eXchange file Format), format plus simple destiné à la manipulation de produits finis ou complets, qui n'offre que des possibilités de montage limité aux « cuts » mais dont le champ d'application est beaucoup plus large puisque visant l'ensemble de la chaîne de production et de diffusion. Il a donné naissance à des normes SMPTE.

Les deux produits comportent de nombreux aspects communs. Ils sont complémentaires et devraient pouvoir être, à terme, interopérables.

MXF a été promu par trois entités : le Pro-MPEG Forum, l'UER (Union Européenne de Radiodiffusion) et l'association AAF.

Les premiers travaux ont été présentés par le Pro-MPEG Forum lors d'une réunion technique en juillet 1999. MXF est un ensemble de normes. Les premières normalisations par le SMPTE (Society of Motion Picture and Television Engineers) ont eu lieu en 2003 mais tous les documents ne sont pas standardisés aujourd'hui. C'est un travail récurrent de normalisation et de révision.



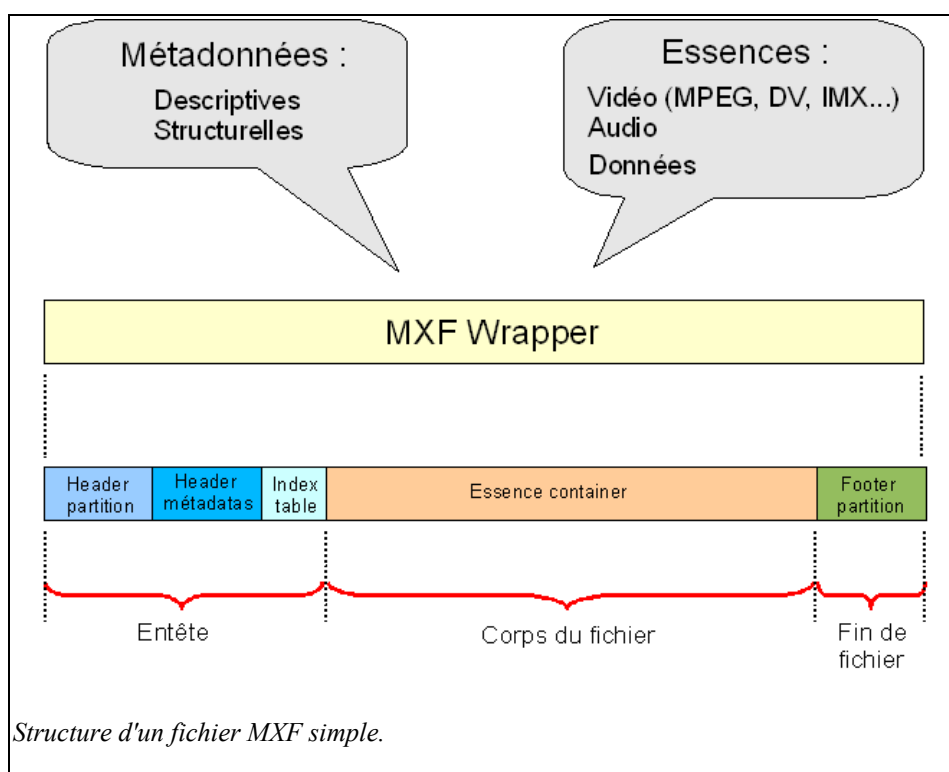
En parallèle aux composantes principales du programme (les images vidéo, les canaux audios et des données) vont donc être générées des données particulières destinées à réaliser cette mission de renseignement. Un nouveau vocable pour désigner ces informations supplémentaires est né : on les appelle métadonnées. Les métadonnées sont donc des données sur les données « utiles » (l'audio, la vidéo et les données), ces dernières étant désormais désignées sous le terme d'essence (la vidéo est une essence, l'audio est une autre essence...). Il ne faut pas confondre données et métadonnées : les premières sont des **éléments constitutifs** du produit audiovisuel au même titre que l'audio et la vidéo (par exemple des sous-titres), les secondes sont des **informations** sur le média audiovisuel.

MXF se distingue par son format de fichier et son mode d'encapsulation particulier des données, par sa gestion spécifique des médias audiovisuels qui repose sur une distinction entre niveau physique (le stockage des contenus) et niveau logique (les instructions de lecture ou d'exploitation), et par l'intégration des métadonnées dans l'enveloppe même du fichier qui, de ce fait, ne seront jamais dissociées des essences correspondantes.

1 - Le fichier MXF : l'encapsulation et l'indexation des essences

MXF est avant tout un format de fichier. C'est un conteneur (wrapper en anglais), c'est à dire qu'il peut intégrer dans une unique enveloppe l'ensemble des données liées à un programme : les composantes vidéo, l'audio, les données et les métadonnées. Cette forme de fichier n'est pas nouvelle : le format AVI (Audio, Video, Interleave) créé en son temps par Microsoft ou QuickTime proposé par Apple sont des exemples de ce type de format « capsule » mais MXF est beaucoup plus complet et ses fonctionnalités plus étendues.

Le format de fichier MXF semble à priori classique (une entête, header en anglais) contenant des informations sur l'ensemble du fichier et intégrant notamment des métadonnées, un corps (body) contenant toutes les données des essences et une fin de fichier (footer). Un examen plus détaillé permet de révéler toutes les spécificités de ce format. Au sein d'un fichier MXF, les données audiovisuelles et les métadonnées sont stockées dans le corps du fichier (body) dans des entités spécifiques, les Generics Containers. Ce sont eux qui assurent le stockage des contenus. Le corps du fichier peut intégrer un seul ou de multiples containers.



Pour les données vidéo, un ensemble de normes décrivant l'implémentation des principaux formats dans ces containers sont aujourd'hui disponibles : SMPTE 381 M pour le format MPEG (incluant MPEG-1 et MPEG-2), SMPTE 383 M pour la famille des formats DV, SMPTE 384 M pour les images vidéo non compressées, SMPTE 385 M pour les flux SDTI-CP, SMPTE 387 M

pour le HDCam (D11)...

Le format MXF est non seulement indépendant du standard mais également des caractéristiques dimensionnelles des images vidéo (525 ou 625 lignes, SDTV ou HDTV...) ou du débit utilisé (de 500 Kb/s à 1,5 Gb/s). MXF est également indépendant vis à vis des types de plateformes informatiques et des systèmes d'exploitation utilisés (Windows, MacOS, Unix, Linux...) ou par rapport aux protocoles réseaux. En dépit de cette souplesse, MXF n'est pas un format de compression : même s'il peut transporter indifféremment les différents formats, il n'assure pas leur transcodage. Pour toutes les données, le débit peut être constant ou variable.

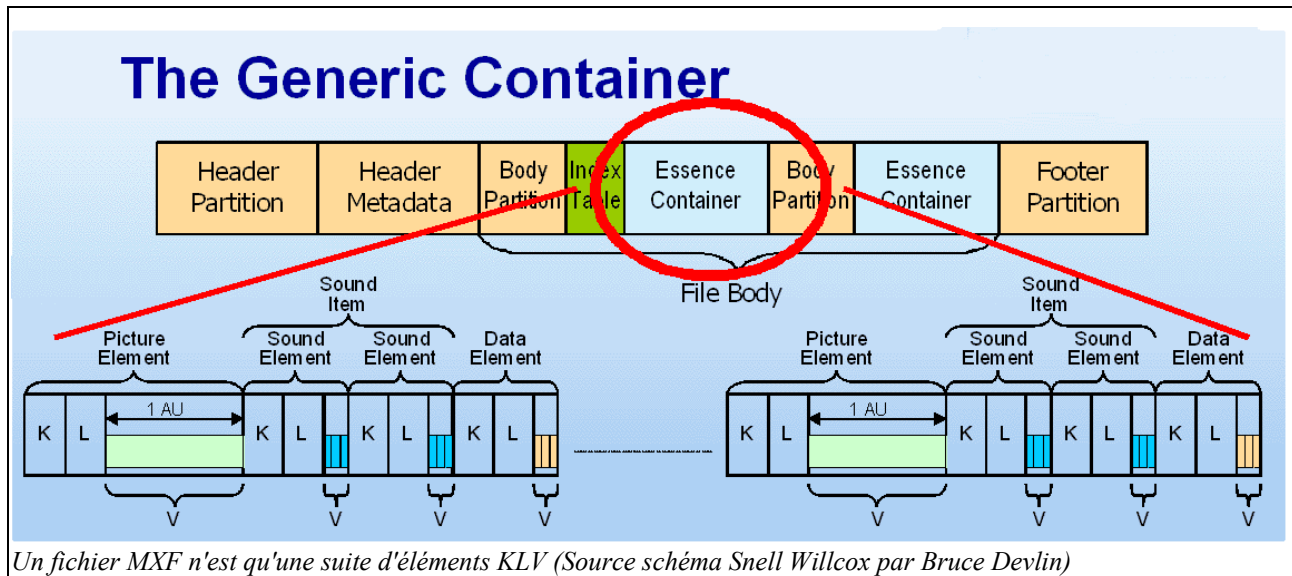
Un Generic Container peut être mono-essence ou « interleaved », c'est à dire intégrer des données audio et vidéo entrelacées offrant alors des fonctionnalités spécifiques comme l'ouverture partielle de fichier ou la diffusion en streaming. Ce dernier point est important. Par rapport à la notion de transfert, un fichier MXF peut en effet offrir deux types de fonctionnement : transfert « classique » de type fichier informatique (par exemple par FTP) où il est nécessaire d'attendre que la fin du fichier soit reçue avant d'en commencer l'exploitation ou transfert par streaming, (sous la forme d'un flux continu donc) où il est possible de décoder et de visualiser les contenus audiovisuels pendant le transfert et avant qu'ils ne soient totalement transmis.

La forme simple de fichier est suffisante dans la plupart des cas. Des formules plus compliquées sont cependant possibles. On peut y ajouter des tables d'index (elles permettent un accès rapide et aléatoire en tout point des essences, mais sont également utiles pour une restauration partielle des fichiers), re-découper chacun des trois ensembles constitutifs du fichier MXF en partitions...

Les paquets KLV : la structure de base du fichier MXF :

Au delà de cette forme globale somme toute classique (une entête, un corps...), un fichier MXF peut être décomposé en une suite d'éléments ayant une structure beaucoup plus fine. Le découpage en paquets réalisé utilise un codage en KLV pour coder chacun des morceaux constitutifs, quel que soit le type de données concerné. Chaque paquet comprend trois champs. Le premier est une clé d'identification unique de 16 octets (K comme Key) concernant le contenu qui pourra être soit du son, soit de la vidéo, soit des données. Le second est une indication de la longueur des données (L comme Length en octets). Le troisième champ contient les données réelles. Les outils, notamment les plus simples, qui ne reconnaissent pas une clé ignoreront purement et simplement l'élément qu'ils ne peuvent interpréter. Cette architecture est une garantie pour l'évolution du format, l'ajout de nouvelles fonctionnalités se traduisant par la création de nouvelles

clés. Au final, un fichier ou un flux MXF n'est donc qu'une longue suite de paquets KLV placés les uns à la suite des autres. KLV est défini par le document SMPTE 336M.



2- La gestion des données audiovisuelles dans le fichier :

Au sein d'un fichier MXF, il y a séparation des contenus (audio, vidéo) et des instructions de lecture les concernant. Les « packages » interviennent dans le traitement des contenus et ils décrivent les manières dont les éléments sont liés et exploités. Trois types de package ont été définis.

- l'ensemble Material Package ou MP définit la manière dont les essences sont lues : il assure la chronologie du déroulement, un peu comme une liste de diffusion mais il ne contient pas de données. Il représente la ligne de temps du fichier MXF (pour chacune des essences ou des données, on y a d'ailleurs défini la notion de piste, track en anglais). MP gère la synchronisation et l'ordre de lecture des portions sélectionnées.
- l'ensemble File Package ou FP représente le stockage physique des essences dans les containers ou même à l'extérieur des fichiers MXF. Il fournit les indications pour la localisation des différentes essences dans le fichier. Il lie les pistes du MP aux containers des essences.
- l'ensemble SP (Source Package) contient des données relatives au contenu (listes de montage précédentes, description...).

Dans le même esprit que pour les normes MPEG, et pour répondre aux caractéristiques et aux contraintes des différentes applications possibles au long de la chaîne de production de diffusion, MXF a introduit la notion de profils. Dix variantes structurelles répondant au nom de OP

(Operational Patterns), plus ou moins complexes suivant les applications visées, captation, montage ou diffusion sont prévues. Comme schématisé dans le tableau ci-après, elles se déclinent par une numérotation de 1 à 3 avec pour chacune d'elle 3 niveaux de complexité repérés par les lettres a , b et c.

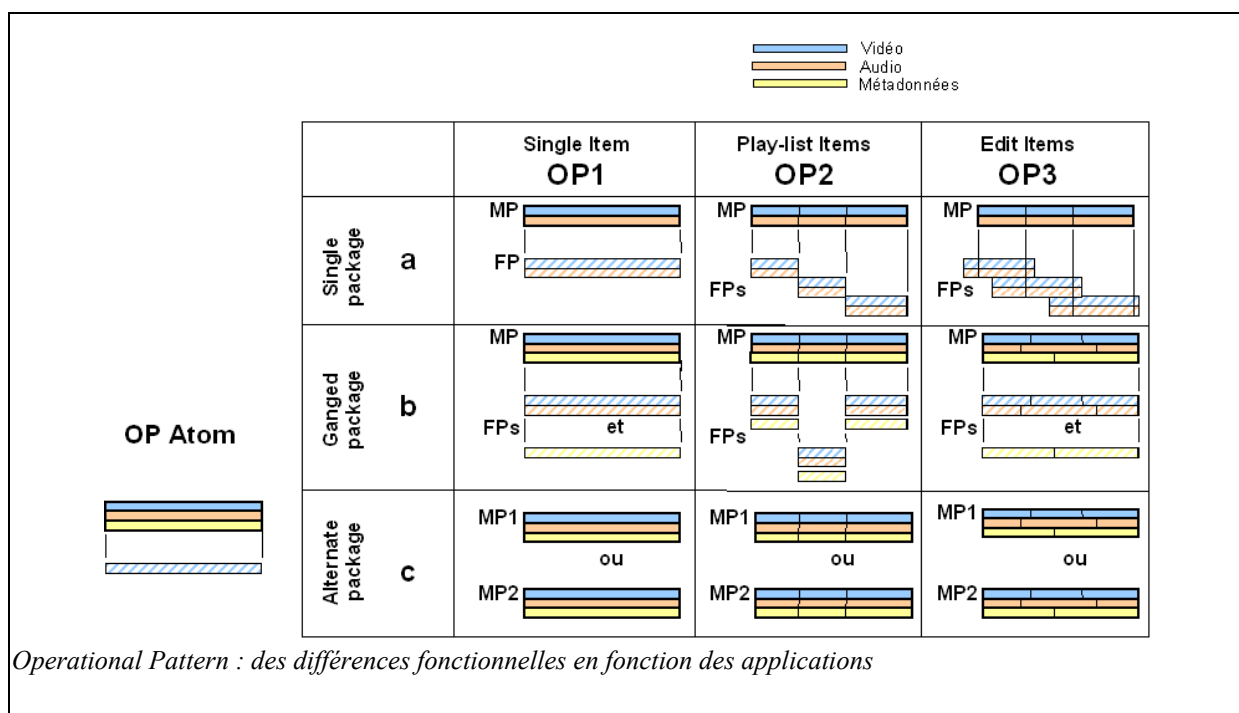
OP1 correspond à la lecture d'un seul container d'essence (une vidéo avec l'audio et les métadonnées associées) dans un unique fichier avec une lecture en totalité.

OP2 vise les applications de diffusion : à la manière d'une play list, elle autorise la lecture d'une succession de containers entiers.

OP3 correspond à des fonctionnalités de montage et permet un adressage, non pas à des containers entiers, mais à des portions de containers (accès aléatoire aux médias).

Pour chacun de ces trois niveaux, trois degrés de complexité donc. Le niveau a est le plus simple. Le niveau b (ganged package) autorise le groupement de plusieurs contenus (FP). Le niveau c (alternate package) indique la possibilité d'exécuter le fichier MXF à partir d'un container ou bien d'un autre (par exemple pour offrir différents contenus linguistiques pour une même séquence vidéo).

La dixième variante structurelle est baptisée OP Atom : c'est le format le plus élémentaire de tous puisqu'il n'intègre qu'une seule et unique essence dans le fichier. Il est particulièrement adapté au montage. Cette forme est à comparer avec le profil OP1a qui intègre dans un même fichier les essences vidéo, audio et données alors que pour OP Atom, ces essences sont dans des fichiers séparés et liés par des UMID (avec le même Material Package).



Dans la réalité, toutes ces combinaisons ne sont pas utilisées. Seules les deux OP les plus basiques sont pour l'instant réellement employées. Ce sont les OP1a (par Sony notamment avec le XDCam) et OP Atom (par Panasonic pour ses produits au format P2). Les serveurs (Omneon, Pinnacle aujourd'hui Avid, SGI...) supportent dans leur majorité le profil OP1a. Omneon supporte récemment le Op1b avec références externes vers les essences. Les MXF avec références externes représentent une tendance forte en ce moment chez les broadcasters. A côté des Serveurs Pinnacle (aujourd'hui Avid) ou SGI, on peut aussi citer Thomson, Seachange et Leitch.

Des passerelles sont possibles entre OP1a et OP Atom. Au sein du SMPTE, on envisage d'ailleurs la création d'un nouveau profil qui reprendrait les points les plus positifs de ces deux profils.

3- Les métadonnées :

Les métadonnées sont donc des données sur les données, c'est à dire des informations sur les essences contenues dans le fichier MXF et qui vont permettre leur exploitation. Elles peuvent être générées à la source, dès la prise de vue par l'intermédiaire des équipements de prise de vues et/ou ajoutées ensuite tout au long de la chaîne de traitement, à la production, pour la diffusion ou pour l'archivage.

Elles sont de deux catégories :

- les métadonnées structurelles : c'est à dire concernant la structure même des essences. Elles définissent la nature même de l'essence ainsi que des paramètres «techniques» (format, taux d'échantillonnage, codecs utilisés...) et sont souvent créées automatiquement avec le fichier. Elles sont nécessaires aux équipements pour l'exploitation des contenus.
- les métadonnées descriptives : ce sont des données informatives et non pas techniques qui répondent à des préoccupations éditoriales ou descriptives. Elles fournissent des informations ou des annotations sur le contenu audiovisuel : titre du document, lieu de la prise de vue, nom des participants, droits, indications du script, évènements... Elles ne sont pas nécessaires pour le décodage des essences et ne participent pas au résultat final. La norme DMS1 (Descriptive Metadata Schema) définit un modèle pour ces métadonnées, mais c'est aussi un dictionnaire (elle décrit une liste de métadonnées et un modèle de hiérarchisation). Elle est encore peu implémentée aujourd'hui.

Au sein d'un fichier MXF, les métadonnées peuvent être localisées en divers points : dans l'entête (header métadatas) et/ou dans le corps du fichier.

L'entête peut contenir des données structurelles ou descriptives. Elle donne des indications sur la façon dont doit être lu le fichier MXF, par exemple sur le format des essences contenues dans les Generic Container (DV, MPEG, non compressé...) ou sur les métadonnées qui sont incluses dans le fichier. Une référence UMID est également intégrée dans l'entête, mais elle est uniquement destinée à l'identification du fichier MXF lui-même. Les équipements de montage pourront y ajouter des informations spécifiques (description de scènes, EDL...).

Au niveau des métadonnées, la forme de concaténation utilisée dans le fichier MXF présente un autre avantage pour les utilisateurs : les métadonnées situées dans l'entête peuvent être analysées même si les vidéos intégrées dans le corps du fichier ne peuvent pas être décodées et exécutées (codec absent dans l'équipement destinataire par exemple). Conséquence, il est tout à fait possible d'exploiter ces informations et d'identifier le produit audiovisuel sans qu'il soit nécessaire de décoder les essences correspondantes.

Dans le corps du fichier, les métadonnées sont intégrées dans les Generic Container au même titre que les données audiovisuelles. Pour reprendre la terminologie employée, un Generic Container peut être divisé en «items» qui sont eux-mêmes divisés en éléments, eux-mêmes codés en KLV. Un Generic Container contient en fait quatre types d'item : picture et sound item, c'est à dire les données vidéo et audio, System Item, le plus important car contenant des paramètres spécifiques pour la gestion des flux et des métadonnées comme le time code, l'UMID... des éléments audiovisuels contenus dans le fichier. Des métadonnées peuvent être également placées dans un quatrième Item : data Items (données non audiovisuelles, sous-titres par exemples).

Lors de la réalisation d'un produit audiovisuel, les premières métadonnées sont généralement générées dès la phase de prise de vue, par les équipements eux-mêmes, notamment celles concernant le temps et le lieu. On peut citer par exemple le time-code et l'UMID.

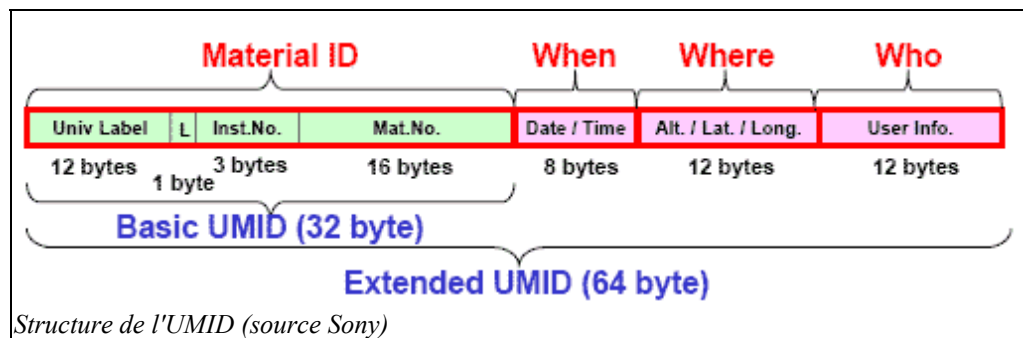
Le time-code est certainement l'un des premiers outils. Né avec les magnétoscopes analogiques, inséré dans le signal vidéo lui-même (VITC Vertical Interval Time Code) ou enregistré sur une piste séparée sur la bande magnétique (LTC Longitudinal Time Code), il n'apportait pas d'information sur un contenu mais fournissait une référence temporelle permettant la localisation d'un point sur la bande avec toute la précision nécessaire au montage, c'est à dire à l'image près. Le repérage de chaque image est réalisé sous la forme d'un codage horaire complété par un numéro

d'image parmi 25. Il s'appuie aujourd'hui sur un code de 80 ou 90 bits. Il est normalisé sous la référence SMPTE12M.

L'UMID (Unique Material Identifier) est plus récent, et est également standardisé par le SMPTE (SMPTE 330M). Il est destiné à l'identification des matériaux audiovisuels. Cette référence est unique. Il est composé de deux parties sur 64 octets :

- l'UMID de base (32 octets) identifie une seule unité de contenu (un clip complet audio/vidéo, un montage) par exemple par introduction du code dans la caméra de prise de vues.

- l'UMID étendu (32 octets) contient des données complémentaires (when, where, who) tels que la latitude, la longitude, l'altitude (les caméscopes modernes peuvent être équipés de GPS), la date, l'heure, l'utilisateur... Cette extension peut identifier chaque partie constitutive.



MXF est une norme jeune et elle commence à être implémentée dans la chaîne audiovisuelle. Plus que dans la post-production, c'est au niveau de la captation d'une part et pour les échanges de produits finalisés d'autre part qu'elle est aujourd'hui davantage utilisée. La mise en oeuvre d'une évolution de cette ampleur nécessite du temps. Souvenons nous : il aura fallu attendre 8 ans pour que toute ambiguïté disparaisse dans l'implémentation de MPEG-2. Défaut de jeunesse, des problèmes d'interopérabilité peuvent encore apparaître, mais MXF est sans aucun doute l'une des évolutions majeures de ces dernières années pour l'audiovisuel.

Quelques unes des normes MXF au SMPTE

SMPTE 377 M (MXF File Format Specification)

SMPTE 378 M (OP1a)

SMPTE 379 M (Generic Container)

SMPTE 380 M (Descriptive Metadata Scheme - 1 (DMS-1))

SMPTE 381 M (Generic Container MPEG)

SMPTE 383 M (Generic Container DV)

SMPTE 384 M (Generic Container Uncompressed)

SMPTE 385 M (Generic Container SDTI)

SMPTE 386 M (Generic Container D10)

SMPTE 387 M (Generic Container D11)

SMPTE 388 M (audio A-Law)

SMPTE 389 M (System Item – Reverse Play)

SMPTE 390 M (OPAtom)

SMPTE 391 M (OP1b)

SMPTE 392 M (OP2a)

SMPTE 393 M (OP2b)

SMPTE 394 M (System Scheme 1 for the MXF Generic Container)

SMPTE 405 M (Elements and Individual Data Items for the MXF Generic Container Systems Scheme 1)

SMPTE 407 M (OP3a et OP3b)

SMPTE 408 M (OP1c, OP2c et OP3c)

SMPTE 422M (Jpeg 2000)

SMPTE 434M (XML Encoding for Metadata and File Structure Information)

SMPTE 436M (Television - MXF Mappings for VBI Lines and Ancillary Data Packet)

SMPTE EG41 M (MXF Engineering guideline)

SMPTE EG42 M (DMS1 Engineering guideline)

D'autres documents sont dans la phase finale de normalisation (Final Committee Draft), par exemple la norme SMPTE 410M (Generic Stream Partition). D'autres en sont encore aux études préliminaires : les travaux viennent ainsi de commencer pour les sous-titres ou pour le format VC-1, pour le Mpeg4 AVC-I, le VC-3 (DnxHD).

Des travaux ont été également entrepris en direction du cinéma numérique.

Le 377M et le EG42M constituent les documents fondamentaux à lire en priorité.