BBC

R&D White Paper

WHP 041

September 2002

The Application of Intimate Metadata in Post Production

P. W. Walland[1], G. Thomas[2], M. Koppetz[3], J. Cardoso[4], T. Erseghe[5],F. Hericourt[6]

[1]Snell & Wilcox Ltd., UK, [2]BBC R&D, UK, [3]Arnold Richter, Germany, [4]INESC Porto, Portugal, [5]University of Padova, Italy, [6]France 2, France

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 041

The Application of Intimate Metadata in Post Production

P. W. Walland[1], G. Thomas[2], M. Koppetz[3], J. Cardoso[4], T. Erseghe[5],F. Hericourt[6]
[1]Snell & Wilcox Ltd., UK, [2]BBC R&D, UK, [3]Arnold Richter, Germany,[4]INESC Porto, Portugal, [5]University of Padova, Italy, [6]France 2, France

Abstract

With the advent of ever more complex metadata schemes to describe content throughout the production and archiving process the possibility is arising to create and use additional information alongside the captured essence which enhances manipulation and content formatting capability in post-production. The creation and use of this "intimate metadata" is set to revolutionise production and post-production operations, and will be a core element of managing the convergence of content gathering and programme creation in IT-based systems.

In this paper a number of forms of intimate metadata will be introduced and their applications and benefits described. Examples will be presented of specific metadata-based production carried out under the IST MetaVision project, ranging from innovative electronic camera implementation, incorporating temporal (motion) and 3-D (depth) metadata; its storage and post-production to a conformed output programme using intimate metadata to drive editing and format conversion processes; and compression and transcoding operations to create a wide range of distribution formats which re-use existing metadata. The architecture and particular challenges of implementing such a production chain will be described. Progress in the MetaVision project and the likely time-scale of results will be presented.

This document was originally published in the Conference Publication of the International Broadcasting Convention (IBC 2002) Amsterdam, 12-17 September 2002.

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

THE APPLICATION OF INTIMATE METADATA IN POST PRODUCTION

P. W. Walland[1], G. Thomas[2], M. Koppetz[3], J. Cardoso[4], T. Erseghe[5], F. Hericourt[6]
[1]Snell & Wilcox Ltd., UK, [2]BBC R&D, UK, [3]Arnold Richter, Germany, [4]INESC Porto, Portugal, [5]University of Padova, Italy, [6]France 2, France

ABSTRACT

With the advent of ever more complex metadata schemes to describe content throughout the production and archiving process the possibility is arising to create and use additional information alongside the captured essence which enhances manipulation and content formatting capability in post-production. The creation and use of this "intimate metadata" is set to revolutionise production and post-production operations, and will be a core element of managing the convergence of content gathering and programme creation in IT-based systems.

In this paper a number of forms of intimate metadata will be introduced and their applications and benefits described. Examples will be presented of specific metadata-based production carried out under the IST MetaVision project, ranging from innovative electronic camera implementation, incorporating temporal (motion) and 3-D (depth) metadata; its storage and post-production to a conformed output programme using intimate metadata to drive editing and format conversion processes; and compression and transcoding operations to create a wide range of distribution formats which re-use existing metadata. The architecture and particular challenges of implementing such a production chain will be described. Progress in the MetaVision project and the likely time-scale of results will be presented.

INTRODUCTION

In this paper an integrated architecture is introduced which has been developed by the MetaVision project based upon using and extending existing metadata standards. The concept of "intimate metadata" is introduced, and it is shown how this can be used in production, post-production and distribution to enhance the value of the content, maintain the original captured quality and support format conversions and post-production operations. The MetaVision format is introduced, based upon high resolution frames at 24 frames per second and lower resolution frames at a higher rate to provide motion information. The implications of the capture data rate on bandwidths and the need for lossless compression are described. The creation of intimate metadata providing depth information is an important aspect of the project, and the way in which this is achieved using a stereo capture system is presented, along with the ways in which this additional information can be used in post-production. The handling of the metadata, referencing and exchange has been investigated in the project, and the MXF format has been

selected, with a CORBA based distributed management system. Finally, the techniques of using intimate metadata in post-production and for formatting and distribution are described, along with the new working methods which are made possible using the MetaVision format.

CAPTURE OF IMAGE CONTENT AND CREATION OF INTIMATE METADATA

The process of capturing an image for later display, no matter what technique is used, is always a compression process. In the real world any scene is immensely complicated and continuously changing, and the amount of information necessary to describe it completely would be beyond the scope of any system imaginable. Therefore, we are forced to restrict the field of view and take short time snap-shots in the hope that the human observer can be fooled into believing that they are seeing a facsimile of the real world. Different techniques have evolved over the years for film based imaging and electronic based imaging, and very often these formats need to be converted one to the other. The most important element in handling any of these images is the avoidance of artefacts in the final display. After all, the intention of any film or camera work is to convey to an audience a sense of presence in the scene portrayed, which involves a suspension of belief on the part of the viewer. Any intrusive artefact in presentation will come between the audience and the artistic intention of the production team. It is not surprising, therefore, that much technical work and ingenuity has been expended on trying to reconcile the limitations of information handling capacity with the need to reproduce an original moving image as faithfully as possible.

It is in this environment that the MetaVision project has been working to take the most recent standards and technology and create a complete image handling chain which brings together the different image creation techniques, and by the use of new media handling formats combines the capture of core "essence" with additional visual information which can reduce artefacts and maintain image quality, Walland [1].

However the image is captured there is necessarily a time-slicing effect which leads to the sort of familiar artefacts such as wagon-wheels turning backwards and helicopter blades "jumping". Fast motion is a particular problem and is normally addressed on film by using motion blur to disguise the effect and by avoiding the kind of camera motion that can accentuate it. In television capture interlace "spreads" the motion across adjacent fields without increasing the effective data rate. In MetaVision one of the additional sources of information, or "intimate metadata" captures the motion in the scene. High resolution images are captured at standard film rate (24 frames per second) and the motion within the scene is captured at a much higher frame rate (72 fps up to 150 fps) but transferred at a lower resolution per frame. This means that the data rate is not unduly high but that high-resolution frames can be reconstructed from a knowledge of the closest frames and the motion between them.

DEPTH METADATA IN METAVISION

When a natural scene is imaged onto a film or an image sensor then the depth dimension is lost and relative positions within the scene can only be inferred by visual cues such as relative size, parallax and occlusion. This depth content is captured in the MetaVision format as stereo images from which a depth map can be generated and associated with the "essence" as further "intimate metadata". Grau et. al [2] and Thomas and Grau [3]. Although it is not the intention within the MetaVision project to create 3-D displays, 3-D information is essential in post-production to allow, for example, convincing integration of real and virtual content and to segment different parts of the scene. Other applications include set pre-visualisation and of course the creation of stereo programmes.

Figure 1 – MetaVision depth sensor

A review of depth sensing methods, as described in [2] shows that there is no single approach that covers all possible production situations. For MetaVision a stereoscopic approach was selected. Stereoscopic systems cover a wide range of applications and possible production situations, in particular this method can be used in a studio environment as well as in an out-door scene. The developed system uses two monochrome auxiliary cameras mounted near the main camera, as shown in figure 1. At the current state of technology such passive methods, although suited to a wide range of application scenarios, cannot necessarily deliver high-quality data at video frame-rate. For most of the applications being considered it is acceptable to store the stereo images and derive the required depth map later, and in the MetaVision architecture the images of the auxiliary cameras are stored together with the images of the main camera and additional metadata. Further metadata from the camera to support subsequent processing, like focus, iris, exposure time and so on are also recorded in MXF (Material Exchange Format, described later in this paper), which ensures that the intimate relationship between stereo images and the essence are retained, and allows the correct processing of the depth information to be performed in post-production, as depicted in figure 2.

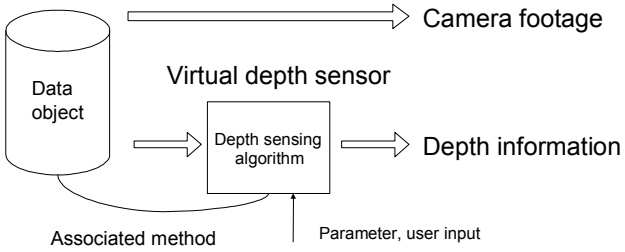

Figure 2 – Use of the data in production

NEED FOR LOSSLESS COMPRESSION IN IMAGE CAPTURE

As time goes on the transmission and recording bandwidth of available devices increases steadily. There is, however, a similar pressure on the bandwidth required driven by the demand for resolutions which are closer and closer to a "film quality" ideal.

Understandably, the creators of original material do not wish to see any compression performed which may cause artefacts to be created later in the production chain. The current "target" capture format for electronic film quality is 4k x 2k pixels, RGB 4:4:4 with a minimum 12 bits resolution at 24 frames per second. This implies a raw data rate of nearly 7 Gb/s. In contrast the current standard rate for HD transmission of 1.48Gb/s represents the transmission rate available for real-time transfer and storage onto disk or tape. For good motion portrayal there is a need to increase the captured frame rate (72 frames per second as a minimum, 150 frames per second preferred) and there is already talk of IMAX resolutions being captured at 8k x 4k pixels. It is fair to say that electronic sensors and cameras which can give this level of performance are not yet available, however it is equally clear that as technology advances on this front so the requirement to transfer and record higher and higher data rates will always be in excess of the available transfer rates.

In the MetaVision project there are two approaches being taken in order to address this apparent contradiction. A level of compression is being achieved by separating the motion content from the image content (as is described in the previous section) and mathematically lossless compression is being developed in order to provide a further reduction in bit rate.

Although lossy compression for video has now reached quite a mature stage, the application of lossless compression to video has not received much attention in the past

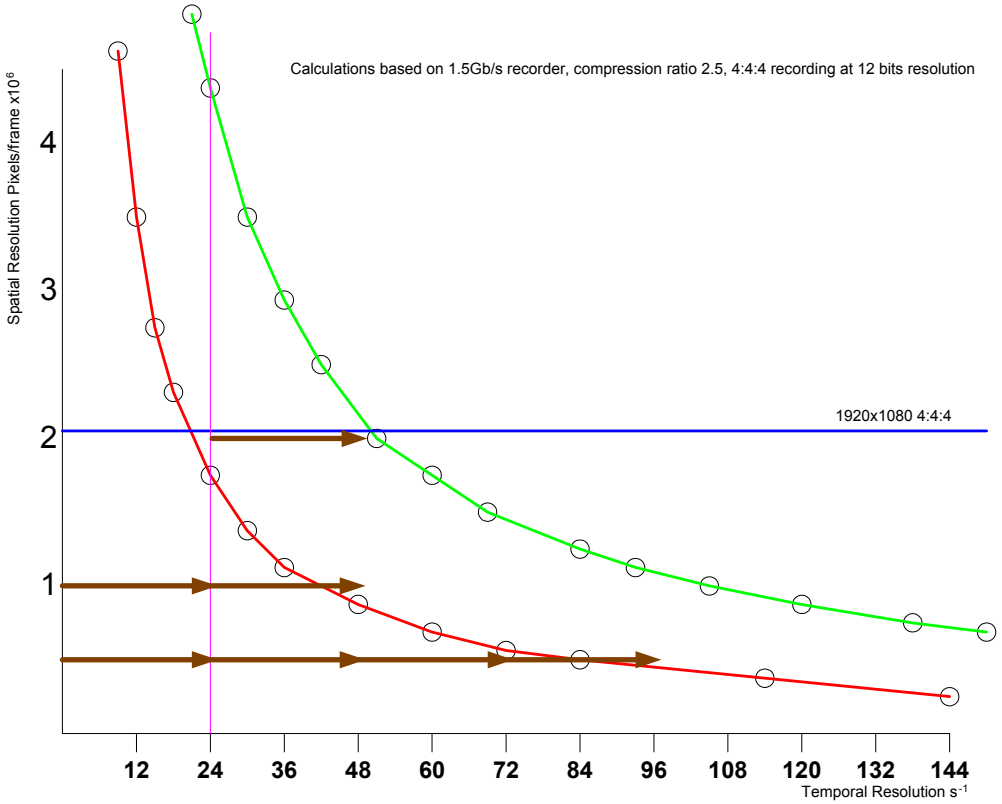

figure 3 - Bandwidth demands as a function of resolution and frame rate

because of the limited compression ratios available. However, with this emerging need for very high quality handling of content in production and post-production it is necessary to develop techniques which will work in this environment.

The graph reproduced as figure 3 shows two curves, the inner (red) curve being the locus of pixel resolution per frame versus transferred frame rate in a standard 1.48Gb/s

bandwidth. (The assumption is that the source is RGB 4:4:4, 12-bit resolution for the high-end quality requirement.) It can be seen from this graph that 24 frames per second capture is on the limit for HD resolutions (2M pixels per frame). If, however, lossless compression is performed giving a compression ratio of, say, 2.5, then there is "spare" bit-rate available for other data. This can be taken up by additional frames at higher rate but lower resolution, as shown by the heavy arrows. As the resolution is reduced so the number of additional frames increases in direct proportion, and it can be seen that at a quarter resolution, a helper signal of up to 96 frames per second can occupy the spare bit-rate. This is the feature exploited by MetaVision, and is only possible when lossless compression is performed on the primary essence.

The work on lossless compression has been undertaken by the University of Padova, who have taken as their baseline the very specific requirements highlighted by the MetaVision project. Any lossless compression performed must be executable in real time, since it is applied to the output of a working camera. This has implications for the maximum complexity of any algorithm selected. It is also essential that the compression algorithm has a controlled degradation path for those occasions when the available lossless compression ratio is insufficient for the connection bandwidth. These two conditions dictate the most appropriate compression algorithm for this activity. A third, and very important element is the need to ensure that the proposed lossless compression method is part of a standardised technique, in order to avoid the creation of a proprietary link in what would otherwise be an open system.

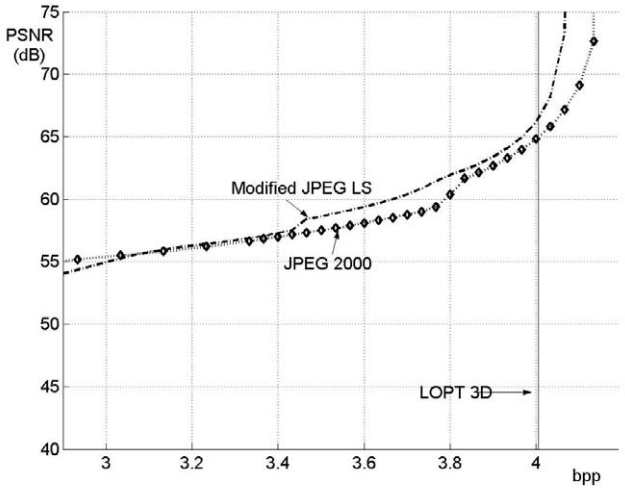

Figure 4 - Performance of lossless compression algorithms under graceful degradation

As can be seen in the graph given as figure 4, it has been necessary to consider the performance of proposed lossless compression schemes under the circumstances in which the target bit-rate cannot be achieved in a purely lossless mode. In this case, the compression scheme must degrade gracefully, preferably in a visually lossless way. LOPT-3D (a lossless coding algorithm using optimal temporal prediction which was developed in MetaVision, Brunello et. al. [4],[5]) is too complex to implement in a mode which permits graceful degradation, and so its performance in lossless mode is shown as an indication of the optimal performance of a lossless scheme at 4 bits/pixel. The asymptotes of both JPEG-2000 and a modified form of JPEG-LS (which allows for lossy compression) show their respective lossless performance. It can be seen from the graph that, for low bit-rates (that is, difficult material or restricted bandwidth) JPEG-2000 is the better scheme, but that for moderate compression beyond the lossless limit, the modified JPEG-LS should be preferred.

USING MXF IN THE METAVISION CHAIN

One task which the MetaVision project must address is the best way to keep the different components of content, essence and metadata together and synchronised within the complete image handling chain. Besides the core essence and the additional visual information (intimate metadata), there is a lot of metadata captured and contributed by the different modules of the system. Given the wide variety of information and formats that this represents, finding the format that enables its seamless integration is not an easy task!

In respect of essence, we live in a world where file and streaming formats coexist, but not peacefully. When we want to exchange data between two modules from different suppliers, things are likely to become complicated and it is likely that they will not understand each other. The current workaround is to reformat the data, leading to an inherent loss of quality and maybe even loss of data.

This creates a quandary. The decision to use equipment from just one supplier, implying a single-vendor environment, will restrict the solutions available and may miss the ability to choose the best product for a given task. On the other hand, mixing equipment from several providers raises the problem of guaranteeing interoperability between them. It would be helpful to have a universally agreed format for content exchange, understood by all. This is where MXF – **M**aterial E**x**change **F**ormat – comes into its own. [6], [7].

MXF aims at universality. The MXF specification is compression format independent and therefore provides for metadata transparency between systems that use different compression formats. MXF is easy to use – an open source MXF **S**oftware **D**evelopment **K**it currently being developed by INESC Porto, flexible and well documented, minimizes the integration time and effort to build MXF compliant applications. MXF is robust against errors – the ability of applications to ignore unrecognized coded data and the repetition of the header metadata provide further reliability. With MXF it is possible to give full rein to the ambition of having components seamlessly integrated in a heterogeneous environment.

These arguments could be refuted by those who see them leading to a potential loss in efficiency. However efficiency was not forgotten in MXF where low implementation overhead and reliability were among the most important requirements. A more detailed look at the MXF format shows how all this was accomplished.

MXF is a wrapper into which a system places video, audio and metadata. Its fundamental purposes are (i) to gather together programme material and related information (both by inclusion and by reference to material stored elsewhere) and (ii) to identify the pieces of information and thus facilitate the placing of information into the wrapper, the retrieval of information from the wrapper, and the management of transactions involving the information.

All the information in an MXF file is coded using Key-Length-Value (KLV) format, defined by the SMPTE 336M standard. To set bounds on the complexity of an implementation, MXF is restricted to 2-level KLV (although the KLV standard allows for infinite nesting). However this does not restrict the logical structure of the file. An MXF file may contain a Logical Grouping of sets or packs that defines a hierarchical structure and where any set may be referenced from any other set. This provides the same logical effect as recursive grouping, and results in all sets being coded in a single layer as a contiguous sequence of sets.

An MXF file comprises three basic elements: header, body and footer:

- The Header gives information about the file as a whole and defines the Template followed by the file. A Template (or Operational Pattern in MXF terminology) determines the complexity of the coder and decoder by defining a subset of the allowed range of parameters. The template identifier in the header indicates what metadata sets are included in the file as well as which of the common essence container formats are used.

- The Body contains the actual audio and video essence, encapsulated in KLV packets.

- The Footer signals the end of the file. Optionally it may include the repetition of the Header Metadata.

To provide an extra level of reliability and enable a decoder to join the transfer after it had started, an MXF file may incorporate File Continuers, which are nothing more than a repetition of the Header metadata. For the same reason MXF decoders must be able to parse any KLV packet and extract the recognized packets whilst ignoring KLV coded data packets that are not known. Upon finding a Key that cannot be determined, decoders should use the length field to skip over the Value of the data packet.

The real strength of MXF in its application to MetaVision comes from its handling of metadata. In an MXF file, metadata is divided into two major categories: structural (describing the internal structure of the essence) and descriptive (e.g. title, producer name, artist and performer). MXF provides a basic metadata descriptive schema – the Geneva scheme. It is a logical structure of metadata sets that allows it to be used as a 'plug-in' to the Header Metadata of an MXF file. This framework and its associated metadata sets may be applied to any MXF Operational Pattern specification.

In the light of the foregoing, MXF becomes the natural choice for the transfer of content between components of the MetaVision chain. A distributed system, CORBA based, that uses MXF for streaming and storage is being developed. Essence and all different types of metadata, including "intimate metadata", are all linked together using the techniques provided by MXF.

Because MetaVision is very much a future looking project, it is inevitable that current proposals and standards are not always sufficiently comprehensive. As the project continues, enhancements will be identified and described. These enhancements and extensions will be submitted to SMPTE for standardisation in due course.

POST-PRODUCTION TECHNIQUES USING THE BENEFIT OF INTIMATE METADATA

One of the advantages of the MetaVision architecture is that the creation, capture and storage of primary information is separated from the output format of the created programme by the various processes and metadata in post production. In principle a wide variety of formats can be accepted and stored, some of which may be compressed whilst other sources may be the full bandwidth of an electronic film camera. The use of MXF as a wrapping medium enables the metadata to be associated with the relevant essence. Moreover, browsing of both audio/visual content and metadata can be performed using the MXF browser and the appropriate decoding tools. The real power of the MetaVision format in post production is that the intimate metadata captured in production can be used to both enhance the editing process and to support the creation of new versions of the

clips in different formats. The following table shows examples of the essence and metadata types which may be handled in post production:

Intimate Metadata	Essence	Metadata	
		Technical	Descriptive
Stereo +	High Resolution	Camera type	Producer
Stereo	High Definition	Focus setting	Location
Depth	Standard Definition	Camera speed	Actors
High Frame rate	MPEG-2	Film type	Crew
Motion	MPEG-4	Lighting	Rights
Etc. …	Etc.	DV	Title
		Etc.	Etc.

The essence is of course the central element of the capture process, but its value is enormously enhanced by the availability of metadata which is closely correlated with the content and can be used in post-production to enhance the value still further.

CONFORMING AND FORMATTING AN OUTPUT PRODUCT

In addition to metadata that is carried in to post production along with the original essence, there will be metadata created at this stage which must remain associated with the essence to which it refers. This is the metadata which refers to how the formatted programme should look, and may also be used to direct the MPEG coding structure which is used. In very simplistic terms, the role of post-production is to create an output product which draws on all the original material available, in whatever format it is stored, identify in- and out-points for each clip and describe the required transition. This is the EDL, which in MetaVision terms is described as a "meta-EDL" because it includes instructions on the required output format of the product; the frame rate, the resolution, the speed, integration of real and virtual effects and so on which define the entire package of content. The meta-EDL will contain references to all the required content using the strength of MXF and unique identifiers to make that reference unambiguously.

MAXIMISING QUALITY/MINIMISING BIT-RATE IN FINAL DISTRIBUTION

The output product needs to be distributed in order to realise its value, and in MetaVision this is at least a two stage process. It is necessary to reduce the effect of artefacts generated by cascading different conversion and encoding processes and this, too, is supported by metadata. The project utilises the re-coding data set standardised by SMPTE (also known as "MOLE") [8], [9] by which the initial coding parameters (first implemented by a master MPEG encoder) are preserved as metadata and re-used for each subsequent encoding, conversion or transcoding operation. The metadata created right at the beginning of the capture process, for example the motion information, can be re-used by the format conversion, rate change and master encoding stages. By using the same metadata at each stage, and then preserving it and re-using as necessary, the very highest quality available can be preserved right through the production to distribution

chain. And because the original material with its original metadata is always preserved for re-use, there is no need to convert between formats with a consequent loss of quality.

NEW WORKING METHODS ENABLED BY METAVISION

MetaVision represents a new way of working which is described in terms of a total working environment and architecture. This new approach enables one to consider new business models for using content and produced programmes. The new workflow is illustrated in figure 5 in which the linear sequence of format conversion and operations which have characterised the handling of content in the past is compared with a new workflow in which there is only one format conversion performed on any piece of content, and metadata is used to support the creation of a conformed output programme. Value is associated with the meta-EDL as well as with the original content clips, and indeed different qualities or formats of clips may be costed differently. Access to archive material is simplified when metadata and MXF transfers are used, thus unlocking the value of such material. The strengths of MetaVision lie in the versatility of the use and handling of the essence and metadata and also in the fact that this is a system concept in which the use of metadata throughout the value chain is considered holistically.

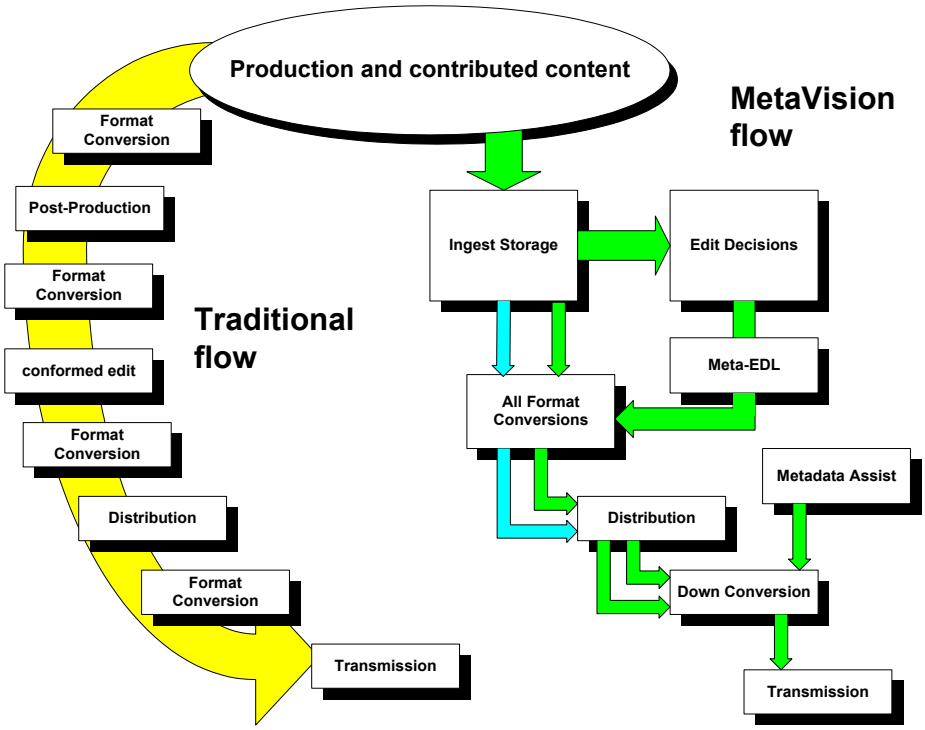

figure 5: workflow for traditional and MetaVision architectures

The MetaVision project is now something over half way through a three year programme. During this time the operating environment and necessary architecture has been investigated and described. The intention of the project is to produce a working demonstrator of the concepts involved in implementing the MetaVision format and metadata supported content handling in an MXF environment. To this end a build programme is underway, which is scheduled for presentation at IBC 2003, following a period of evaluation. Further details on progress are available from the project web-site [10].

ACKNOWLEDGEMENTS

This work was carried out within the European Project IST-1999-20859 "MetaVision".

REFERENCES

1. Walland, P. W., 2002. Metadata for Interactive TV: the challenge of end to end delivery, SAMBITS Workshop "The Converging Environment of Interactive TV and Internet", Munich, March 2002

2. Grau, O., Minelly, S. and Thomas, G.A., 2001. Applications of Depth Metadata, Proc. Of International Broadcasting Convention, Amsterdam, September 2001

3. Thomas, G.A., Grau, O., 2002. 3D Image Sequence Acquisitions for TV & Film Production, Proc of 1st Int. Sym. On 3d Data Processing Visualization and Transmission (3DPVT 2002), Padova, Italy, Jun 19-21, 2002

4. Brunello, D., Calvagno, G., Mian, G. A., Rinaldo, R., 2002. Lossless Video Coding Using Optimal 3D Prediction," to be published in the Proc. of the 2002 IEEE International Conference on Image Processing, Rochester NY, Sept. 2002.

5. Brunello, D. Calvagno, G. Mian, G.A. Rinaldo, R., 2002 Lossless Compression of Video Using Temporal Information, submitted to IEEE Trans. on Image Processing

6. http://www.g-fors.com

7. http://www.pro-mpeg.org

8. SMPTE 319M 2000 - Transporting MPEG-2 re-coding information through 4:2:2 component digital interfaces

9. SMPTE 327M 2000 - MPEG-2 Video Recording Data Set

10. http://www.ist-metavision.com