Joint Task Force on Networked Media (JT-NM) WebFirst UHD Sport Scenario Analysis

22 August, 2016

For more info, go to jt-nm.org

Ownership and copyright

This work is jointly and severally owned by the European Broadcasting Union (EBU), the Society of Motion Picture and Television Engineers (SMPTE), the Video Services Forum (VSF) and the Advanced Media Workflow Association (AMWA) and is licensed under the Creative Commons Attribution-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nd/3.0/ or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Requests for waivers to allow you to use this work in derivative works should be sent to jt-nm-info@videoservicesforum.org

Executive Summary

This WebFirst UHD Sports Scenario was developed as a follow-on activity of the Joint Task Force on Networked Media (JT-NM). The JT-NM is sponsored by the Advanced Media Workflow Association (AMWA), the European Broadcasting Union (EBU), the Society of Motion Picture and Television Engineers (SMPTE), and the Video Services Forum (VSF). The JT-NM was formed to assist the professional media industry in the transition from traditional SDI-based technologies to network-based technologies, with a focus on identifying key areas where a high degree of interoperability is desirable.

Previous activities of the JT-NM have included collection of user requirements, and development and publication of the JT-NM Reference Architecture (RA). More information on JT-NM activities and JT-NM publications may be found at jt-nm.org.

Aiding in a smooth transition to networked-based technologies has always been a goal of the JT-NM. However, from the beginning it was known that enabling the creation of a simple "SDI-replacement using IP" was not sufficient. The JT-NM must look beyond current use cases if it is to produce output that is useful for the industry, even over the relatively short period of the next few years. For that reason, the JT-NM has produced this more forward-looking scenario, which seeks to lay out a set of functionalities and user requirements which will be required in the mid-term, say over the next two to five years.

The JT-NM chose the term WebFirst to draw a contrast with the current situation where many professional media organizations think of the Web as a secondary or alternate method of content distribution and monetization. In the WebFirst scenario, the web is seen as either the primary method of distribution, or at least equal in importance to current methods of distribution.

Another reason the JT-NM wanted to undertake an analysis of a WebFirst scenario was to develop a set of user requirements based on introducing two potentially disruptive concepts to traditional broadcasting; two-way connectivity with the end consumer, and the introduction of Internet Technology and all that brings with it, including big data techniques, large-scale virtualization, and Artificial Intelligence (AI) technologies.

Finally, the JT-NM wanted a way to test the JT-NM Reference Architecture to see if it met a set of user requirements that went beyond a simple SDI-replacement scenario. The thinking is that development of this scenario followed by a gap analysis will allow the JT-NM to determine if there are any fundamental missing pieces in the RA, and let us fill those gaps now before large amounts of infrastructure are developed and deployed.

The JT-NM sponsors hope that readers find that this scenario introduces some new ways of thinking about professional media applications. It does this in an environment that allows us to reuse and monetize our content in different and engaging ways with consumers who expect more from their viewing experience than an SDI-only infrastructure allows.

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

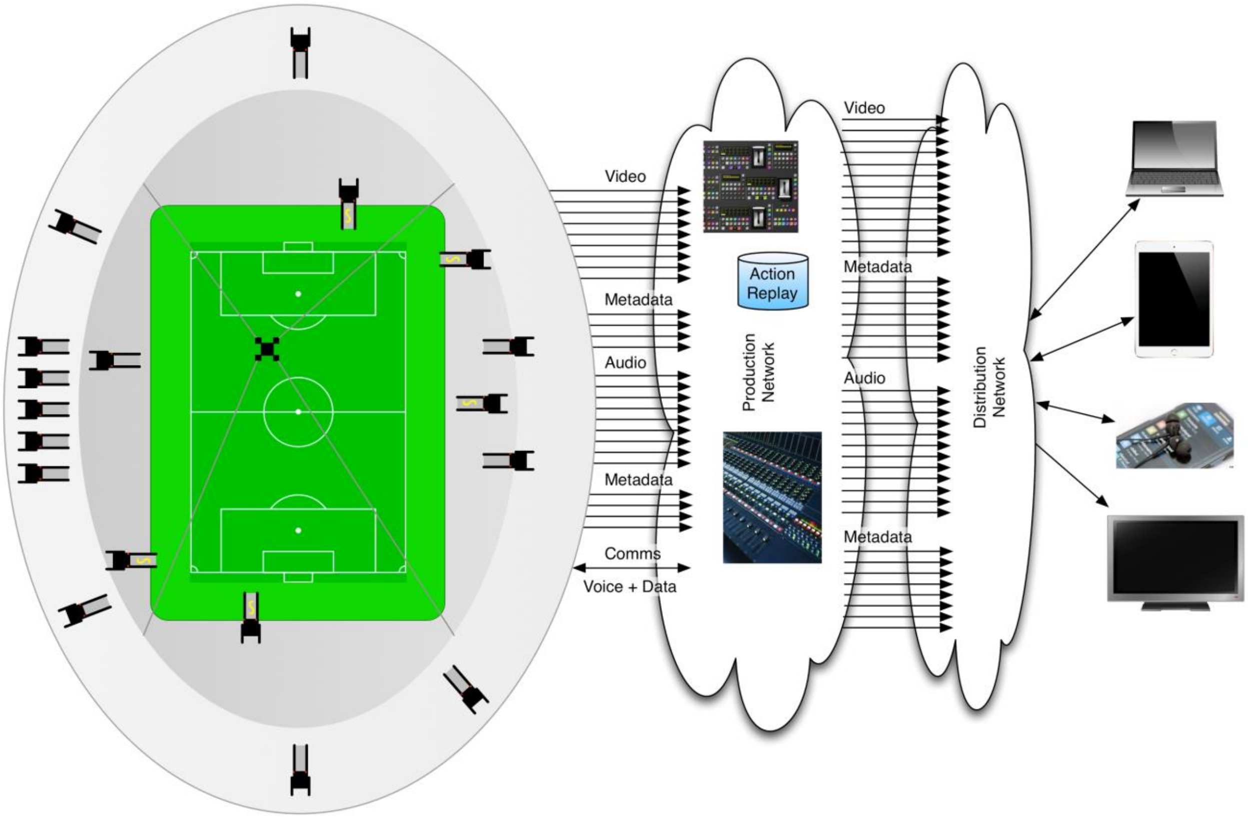
Contents

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

WebFirst UHD Sport Scenario Analysis

The starting point for this requirements analysis activity was a scenario designed to be representative of a future sports production, providing live assets for simultaneous use in a variety of contexts, delivered via multiple distribution technologies to a variety of platforms with widely differing capabilities. This scenario is outlined below:

Scenario



Type of production

High-quality production of a live sports event includes live coverage from multiple stadiums, time-shifted coverage (“as live”), highlights packages, interviews and other features. Audiences can view the event on televisions, web browsers and mobile devices. Audiences will engage differently based upon the device they are using and their location/environment, leveraging the media and time-based information in different ways.

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

Acquisition

Several (typically ten or more) stadium cameras are used to capture the main action of a sporting event. These cameras may have UHD-1 (4k) resolution, frame rates up to 120fps and support high dynamic range. Remote control of camera settings for consistency and convenience is likely to be a requirement. Additional special cameras are used, for example for fixed panoramic “beauty” shots, slow motion replays, close-ups e.g. behind a football net, and 360 degree views. These often have different resolutions and rates and may require additional remote control. Many microphones will be located around the venue.

As well as video and audio, live time-related data is captured, for example camera GPS position, position within the venue, orientation, and zoom. Camera and microphone positions are often changed during a lengthy event, e.g. to reflect the different competitions during an athletics meeting. Additionally, each sport has its own data workflow generally organised around a permanent stream of live data messages (e.g. as results evolve).

Before the event a 3D model of objects of interest in each venue is captured. This data is used to inform substitution of elements within the video such as billboard displays.

Before the event, metadata is provided on the event itself (competition events and sub-events, ceremonies…), the athletes/participants/competitors, teams, delegations, judges, officials, the location(s), start lists, etc.

Live production operations

Live operations on the content include: vision mixing (mostly cuts, with occasional other effects, especially dissolves), metadata for graphics overlay as far as possible downstream (captions, scores, etc.), action replays (including slow motion), audio mixing (for 3D, conventional surround and stereo target systems), addition of commentary and music. All actions are recorded against time (EDL, audio mix parameter adjustments etc.) so they can be used to reconstitute and create new variants of the output for different platforms.

Billboard displays around the stadium are replaced in the video with virtual billboards, enabling the insertion of different advertising for different geographical territories.

Production staff perform live logging via a variety of mechanisms to identify events that happen in the venue, including names of competitors and their location, ‘validity’ of video or audio from each camera or group of microphones. This activity may be assisted or even be performed by automatic AV analysis and/or use of schedule information and live data from various types of sensors.

The time-related data produced by these operations is used to aid editing, and may also be made available in an appropriate way to drive interactive audience experiences. For example a viewer might choose to be notified when a goal is scored or a new leader emerges in a field event, or when a particular athlete is about to perform. A map of the Olympic site could be overlaid with animated hotspots to indicate where interesting eventualities are occurring to aid end user navigation of the content. The desire to be notified of specific events during the coverage such as goals may also drive the overlay of different promo or sponsorship messages or ads.

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

Rendering of audio and video elements can be performed anywhere in the production / distribution chain right up to the end user's device. In the latter case rendering can even be user targeted.

Time-shifting and editing

The production staff can time-shift parts of the event and present them "as-live" within the main coverage, or as action replays of what was already shown.

Edited packages of the event are created for different purposes. Examples include: a traditional highlights programme giving an overview of the event, and shorter highlights packages aimed at web and mobile viewing (e.g. just the goals from a football match, or a particular athletics race). These packages are likely to include different editorial and technical characteristics, e.g. shots may be reframed to suit a different screen, and different audio may be used. The packages will be annotated as appropriate with data derived from logging. The packages can also be created in just a virtual form with the same raw video and audio for all packages.

Sport often requires fast turn-round edits for action replays, often including several versions of the same event from different cameras, including slow-motion cameras. Editing staff will require immediate access to incoming content so they can start working on packages while the event is ongoing.

Production communications

Production staff will use talkback, tally and instant messaging for communications. These can be routed automatically to match the setup of the production.

Staff has access to real-time production information, such as schedules and running orders, regardless of where they are working. This data is linked to live data from other sources (as described previously) allowing complex queries to aid in retrieval of previously acquired content and metadata itself.

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

Requirements Analysis

Analysis was performed through discussion, examining various aspects of the scenario to draw out technical requirements for a system capable of satisfying the scenario.

Scope and Terminology

The purpose of this activity is to think beyond traditional TV and radio as we know it today. For this reason the term End User Experience (EUX) is introduced as a generic term to refer to a broad range of content formats targeted at a wide variety of platforms.

Audio

The system must support a flexible number of channels of audio, without artificial constraints. This is necessary to fulfil the user requirement for rendering on-the-fly at any point in the chain. One approach that could be appropriate in this context is known as "object-based" audio, which requires the tandem carriage of audio pertaining to an "object" (sound source) along with time-varying parameters such as the object's spatial position within the capture environment. This metadata may be modified through the production process and is used to drive rendering to a given speaker configuration. In certain situations this metadata could be generated or modified according to information gleaned from video analysis or by other means, to provide a context-sensitive audio mix to match what's in shot.

Video

The system must be able to simultaneously carry video in different encodings and resolutions, providing formats suitable for different client platforms. Simultaneous working at different video frame rates (and potentially variable frame rates) should also be supported. As well as feeds from live cameras at different frame rates, this may include the mixing of archive material (e.g. 24fps film footage) with video at various rates. Using similar approaches to "object-based" audio, video scenes may be composed of elements that are flexibly rendered in different versions for different target devices or audience segments. These graphical "objects" may be full frames of video, partial frames or more abstractly-described visual elements (3D models). Ultimately these may be delivered directly to the consumer device and the presentation rendered according to preferences set by the end user.

The production system must support simultaneous use of multiple cameras. The working group identified the following (non-exhaustive) list of variants, some of which may be manifest in combination:

- HD
- UHD-1 (4k)
- beyond UHD-1 (> 4k)
- high frame rate
    - higher base frame rate than current max of 50/60p
    - high frame rate capture for slo-mo
    - variable frame rate

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

- high dynamic range
- beauty (fixed panoramic) cameras
- lightfield cameras (perhaps not yet, but in the future)
- 360 degree view
- point of view cameras
- camera drones
- wire cameras
    - spidercam / skycam
    - track cameras
- aerial cameras

Any or all of these variants may be controlled and positioned remotely to provide a variety of shot types from different perspectives.

In addition to the cameras and microphones under the direct control of the professional production crew there may be multiple mobile devices with cameras and microphones in use by members of the audience at the event to capture the action. These devices could be streaming captured content live over the internet via a service such as Periscope or Meerkat, or uploading clips to social media. This provides an alternative source of content that could be tapped by the production team if appropriate, although logistical, quality control and legal considerations may make this problematic.

Data

Data must now be considered to have its own workflow, independent of the traditional media workflow. This is particularly true for sport, involving multiple third parties with different complementary roles. The ability of IP-based media production and distribution systems to carry arbitrary data alongside what we have traditionally regarded as media on the same infrastructure is a key benefit of such systems over the prevailing technology used for live media production. This technique offers opportunities for deep integration of the data workflow into the media production workflow.

For this scenario, data was divided into two separate categories: time-varying, and static.

Time-varying Data

Time-varying data can be gathered from a multiplicity of sources throughout the production chain. This data may be used for various purposes in the course of production, and in some cases may be delivered to the end user / audience member's device to drive aspects of the EUX. Several sub-categories were proposed by the group:

Results

For certain types of sport, data from various measurement systems routinely provide crucial input to the scoring of competitors. These data range from timings (finish times, split times) to distance measurements (e.g. for athletics field events). Of course the final results themselves are of paramount interest to the audience, but the raw statistics can also be used as inputs to the EUX to enhance the viewer's understanding of the event.

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

Telemetry

Measurement of environmental or physiological parameters over time can provide additional insight into the mechanics of sporting performance, which can be of direct interest to audiences. These data may form part of the EUX through visualisation and/or presentation of statistical information. Rendering of these components of the EUX may be performed at any stage between production and presentation, offering progressively greater opportunities for bespoke experiences as the rendering is pushed closer to the end user. Telemetry data may be combined with spatial models to construct graphical overlays.

Some general examples of telemetry sources were identified by the group:

- GPS position of competitors
- speed measurements / accelerometers
- heart rate monitoring / other physiological sensing
- sensors on sports equipment/environmental

Other examples of possible telemetry sources in specific sports were suggested

- Sailing
    - boat position on course
    - angle of heel of boat
    - boat velocity
    - wind speed/direction on course
    - wind as experienced by boat
    - tidal flow across course
- Cycling
    - pedal cadence
    - speed
    - position on course
- Formula 1
    - position
    - speed
    - acceleration
    - engine/braking system parameters (revs etc)
    - fuel level

As well as tracking the subjects of the content and aspects of their environment there is potential value in gathering time-related data from the devices that are used to capture the content (i.e. cameras, microphones and other transducers). The group identified the following parameters:

- Cameras
    - location
        - GPS
        - other location sensing methods
    - orientation
        - focal point (where it’s pointing, where it’s focused)
        - other positional parameters (pose / Az-El etc.)
        - focal length (infer field of view)
        - motion vectors from camera movement

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

- Microphones
    - location (GPS / other location tracking)
    - Azimuth / elevation

The most obvious use for time-varying telemetry data is perhaps to drive graphical overlays that directly represent measurements reported by, for example, a Formula 1 car as it is driven around the track, to enhance the understanding of the viewer. Group discussion around telematics data also surfaced some more subtle applications, particularly where different data feeds may be aggregated and used to guide, inform or automate production decisions:

- video object tracking and/or independent tracking of location of - e.g. cars on a track - could be used to infer origin position of audio sources (audio objects) associated with the tracked physical/video object. This position may be modified subsequently during mixing.

- audio level + location + time used to infer locations/moments of interest, for example in a multi-venue event, the level of crowd noise may correlate with “interestingness”

Since these devices are part of the production infrastructure, in many cases communication links to them can be utilised for bidirectionality, not just reading back parameters but controlling their position and configuration remotely. In turn, dynamic automation of these parameters becomes a possibility.

- device parameters (Read / Write)
    - camera control
        - iris, black level (sit), gain
        - lens: focus / zoom
        - pan/tilt/roll
        - height
        - position (potentially x/y/z, practically often just x or x/y)
        - current, 1st and 2nd derivative values
        - shutter angle
        - shutter type
        - effects wheels
        - ND filtering
    - show control
        - lighting
        - scoreboards
        - teleprompter
    - microphone control

A special category of time-varying data that can be instrumental in enhancing the level of immersion of an experience is tactile or haptic vibration, movement or motion. This can be derived from sensing and/or through synthesis and incorporated into EUXs. There is work being carried out in SMPTE to define standards for encoding and transmission of such data (10E).

Comms

Communication between production staff and between director / producer and presenters, is a critical component of a live production operation. Conventionally this takes the form of audio

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

talkback (intercom) channels used by members of the production team, and Interruptible Foldback (IFB), which is generally fed to presenters via in-ear monitors. With a more flexible IP-based infrastructure at the heart of the system these facilities could be supplemented by text-based communication methods in some circumstances. There may be some value in capturing and storing talkback and other comms alongside the media as another potential source of production metadata (either directly or through some real-time or offline analysis). Some of the information that is traditionally delivered over talkback could be delivered and/or presented in more helpful ways that are sourced from, and linked into production planning and schedule data.

Production Data

Another class of time-related data is that which describes aspects of the production and/or documents decisions made in the production process. This may be generated manually or could be automatically generated through remote sensing or analysis of video and audio. It could be generated in real time along with the live capture of content or added after the fact by logging or post-processing. Some of this data may be specifically related to Quality Control (QC).

Perhaps the simplest way of capturing information about the production is to record it manually.

Logging may relate to what is happening in the event or content captured by a specific camera (e.g. which athlete is featured, when goals occur in a soccer match, occurrences that may be of particular interest for highlights etc.) or may be comments related to the production process (production team members' opinion on usability of acquired media etc.) Traditionally this was done on paper, against a time reference such as SMPTE timecode, but increasingly there are computer-based applications available to streamline this process. IP-based systems offer an opportunity to fully integrate production logging as an additional data feed that can be transported on the same infrastructure as the media and in some cases can be machine-interpreted. Applications built on this infrastructure can readily access both media and relevant pre-existing data, enabling the design of smart tools that simplify the process of production logging.

As well as interpreting logging data, machines can be used to generate logging through analysis of video, audio or live sensing data. Some examples might be voice (speaker) recognition or face recognition used to generate logging events identifying contributors, or speech to text used to generate transcripts. Automated analysis of media can also be applied to gain quality control metrics, detecting, for example, black frames, out of focus pictures or over/underexposure. Flash tests for photosensitive epilepsy are another form of analysis that could produce data marking sections of video as non-compliant.

Crucially, it should be possible to feed any of this information forward for use or further analysis in production processes further down the production chain. In some cases it may be appropriate to use this data directly to drive elements of the EUX. In other cases it may be used in direct or aggregated form as a control input to downstream production operations.

More general logistical and organisational data is used in planning and running the production moment to moment, but can also be useful to supplement and 'make sense of' production data and media elements after the fact. One example may be the combined use of aggregated production and logistical data to inform edit decisions for highlights packages, by automatically identifying and ranking candidate 'clips' by level of interest. The following list was identified by

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

the group as logistical and organisational data that may be of potential value in the production process.

- schedule (when things were supposed to happen)
  - event schedule e.g. order and timing of competitors in a sequential competition such as javelin, gymnastics etc.
  - including location of events (venues, within venues)
  - production schedule (dependent on event schedule)
- when things actually happened / what actually happened
  - record of schedule changes and delays
  - results of individual races / competitions
  - medal tables

Static Data

Certain classes of data are not related to time; these are known as static data. These data may be related to a particular production or event, to configuration of the system at various levels or to aspects of a particular EUX or set of experiences delivered by a production process, independent of time. The following list of types of static data was proposed by the group:

- stable identifiers for competitors, presenters etc.
- stable identifiers for devices / entities in the production system
- 3D model of environment (e.g. stadium or venue)
- static metadata relating to production/event
  - title
  - description
  - categorisation
  - external asset numbers
  - QC test results relating to overall production/event

In the context of static data there was some discussion in the group of web-based semantic linking techniques, allowing navigation through a web of conceptually-similar content either as part of the EUX or external to it. Semantic tagging of content is a relatively low-overhead way of adding machine-readable information that aids reuse in other contexts.

The use of semantic technologies requires a well-defined scheme of identifiers for all objects represented in the data model. The identification strategy may vary from sport to sport. A good predefined identification scheme allows developing independent data feeds on the fly.

For example, the EBU EBUSport ontology covers more than 30 sports and is fully compliant with IOC's ODF data feeds. It is compliant with any other data feed format provided the appropriate data transformation.

Time

Particularly in a live sports context, relationship to time is an important factor in the composition of EUXs from available elements. The progression of real time is one of the only constants in a complex set of equations governing the changing interrelationships of essence and time-related data through the production process. Relationships of essence to time are manipulated in

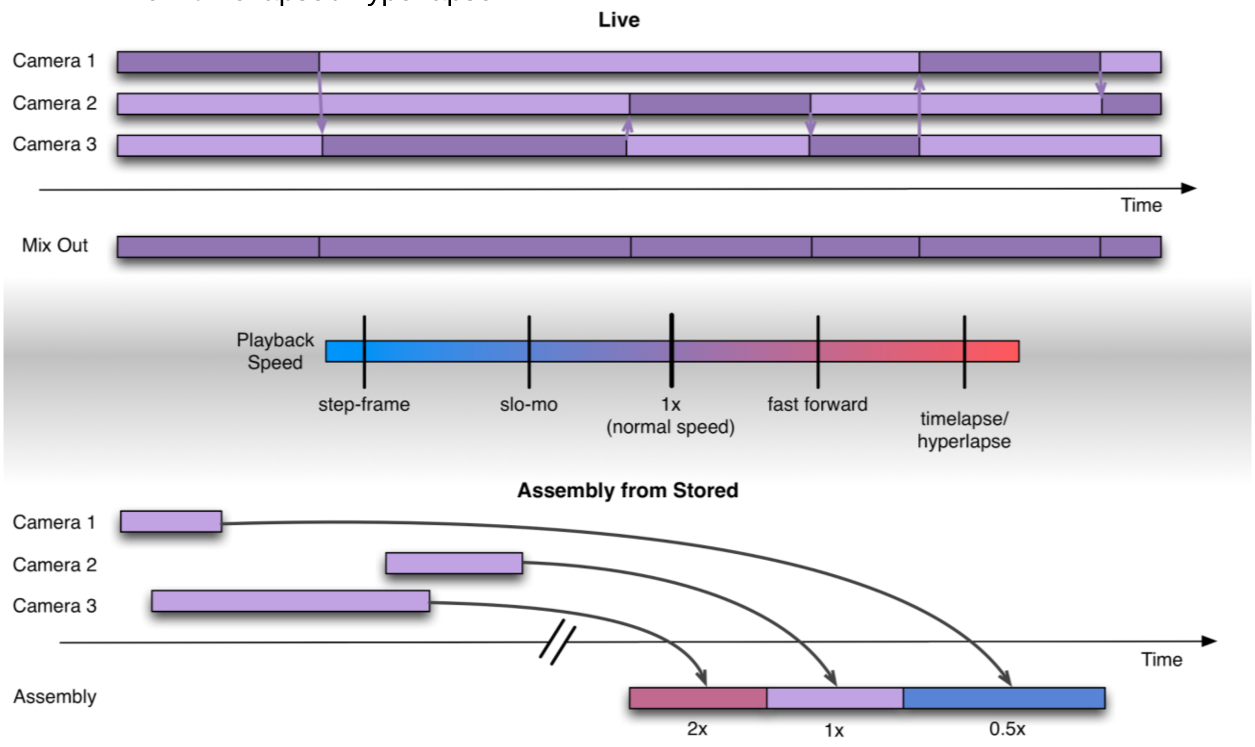
This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

various ways, and these manipulations are dependent upon accurate recording of time at the point of acquisition.

Traditionally, live broadcast and pre-recorded content are treated differently. Workflows and tools have evolved separately, so there is a disconnect between the two worlds. However, sports coverage increasingly uses a mix of live and pre-recorded content. The obvious use case here is simple action replay, but more complex compositions such as highlights packages are also required with faster and faster turnaround. Immediacy has always been linked to value in this context, but parallel delivery to multiple platforms places additional demands on a production team to supply extra content and data streams in real time as well as providing summaries and bite-sized AV clips with minimal delay from the live timeline.

The following list and diagram summarize the various parameters relating to time

- Live
  - direct feed from one or more cameras
  - composition created from cutting/mixing between cameras
- Retrieve/playback from storage (aka time shift from original capture)
  - composition created from assembly of assets taken from different sources
  - slow motion (playback rate slower than capture)
  - normal replay (playback rate same as capture)
  - playback rate faster than capture
  - time lapse / hyperlapse



An ideal system would support the full spectrum between production of live broadcast/streamed content and heavily post-produced content using a common approach for referencing, indexing and synchronisation of content.

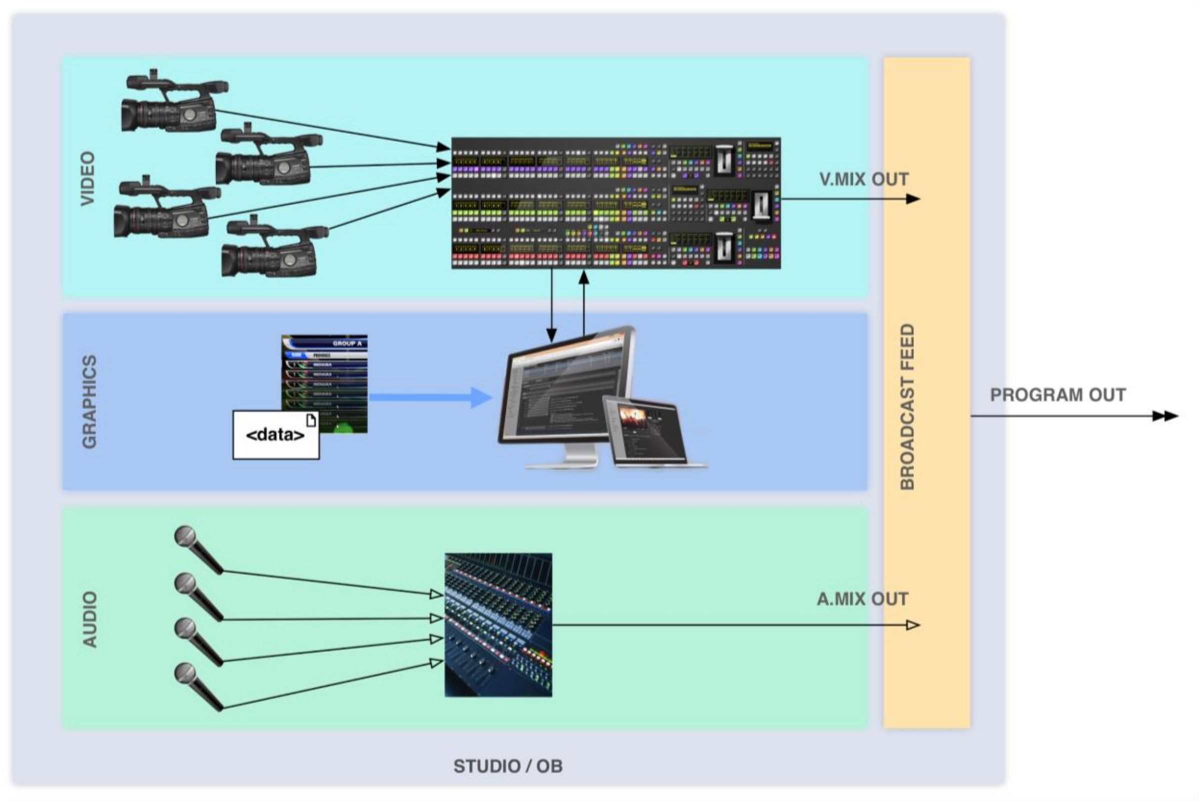
(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

Turning Elements into an Experience for the Audience

Turning video, audio and data feeds into a compelling experience for the viewer is a complex craft, carried out by teams of highly skilled staff and requiring choreographic levels of coordination to produce a single broadcast feed. Our use case requires parallel production of multiple EUXs, each of which may offer a high degree of adaptation or personalisation. What impact does this have on the production process, the tools and the personnel required?

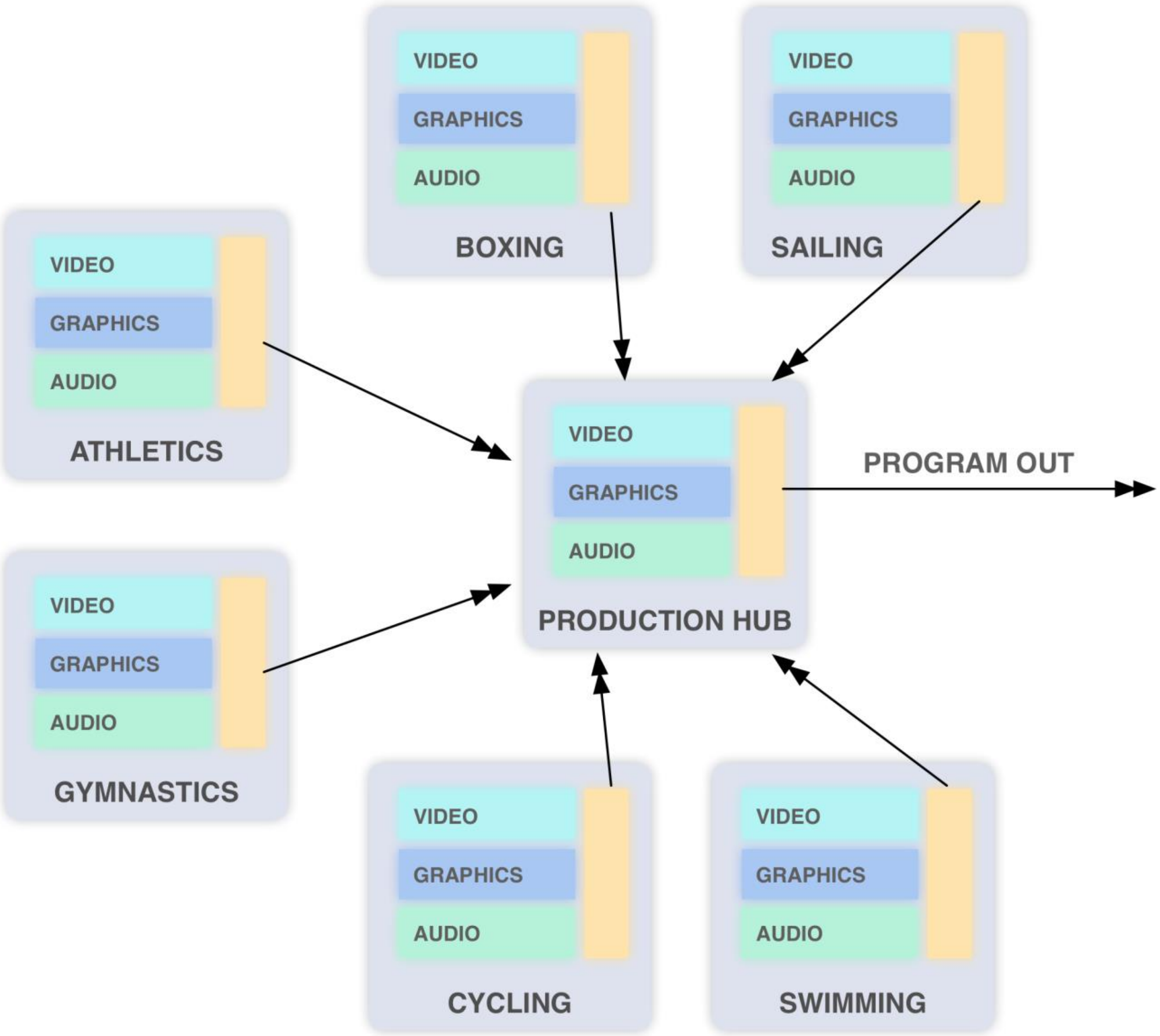
Production operations

Video, audio and graphics for a live show are generally produced by separate teams that coordinate their work so it can be combined to form the broadcast output. While elements such as graphics templates can be prepared ahead of time, any function that relies on live video, audio or data feeds must be performed in real time. The diagram below illustrates a simple production chain culminating in a live broadcast feed:



For coverage of sports events running concurrently in multiple venues, a hierarchy is created, with feeds produced from OB trucks at each venue transmitted to a central production base where they are curated into a single feed for each broadcast channel, connected with live studio presentation spots (note that not all OBs will necessarily need to perform all functions).

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported



The “funnelling” of media through this chain is an approach born out of constraints on communications links from the OB locations to the central production base. It reduces complexity by “sub-mixing” feeds for contribution to the overall production, but it limits flexibility by removing options at the submix stage that can never be reinstated. In a WebFirst production the options filtered out by the OB production team may be required for other purposes; for example to supply synchronised ancillary content to be delivered to a tablet.

Where bandwidth between sites is sufficient it may be feasible to make all source feeds available along with real time control metadata describing any production decisions taken at the OB location. This approach preserves flexibility whilst also communicating the local sub-mixing decisions so that a mix from the OB can be reconstituted from the source elements back at base, where required. Until it’s rendered, the definition of the experience is left open and malleable, so it can be progressively enhanced and/or modified either in series or parallel up to that point. This leads to remote production models that allow content to be sourced from different

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

geographical locations, with the compilation, assembly and composition of the content into EUXs occurring in a variety of (different) locations, with a less rigid organisational and technical architecture. The source elements can be reused in the construction of experiences for different platforms and/or sections of the audience by applying different versions of the composition control metadata at the render stage.

Even where dynamic control automation is used currently, its applicability is generally bounded by the stage of the production process in which it was originated. It rarely transcends these boundaries for various reasons; one of the primary factors is the use of different tools in different stages with no common control ‘language’ enabling interoperability for both live and post workflows. Tools have evolved separately to use different schemas for control data tailored to their specific modes of operation. Distributed production requires that instructions for composition of media from live or stored content are more universally applicable throughout the chain. Moreover these metadata representations need to be live-streamable with similar latency to the media to which they refer.

The following examples of the types of production decisions that may be captured and streamed were suggested by the group. This is a non-exhaustive list, but it at least gives a flavour.

As suggested by one of the group’s members, trying to define a data model to describe everything is something of a fool’s errand, as even if it is possible to cover all current eventualities, there are bound to be things that aren’t yet in our vocabulary. Any standard way of describing composition operations must therefore be extensible.

- vision mixing
    - cuts
    - complex transitions
- audio mixing
    - ambient sound
    - music
    - commentary
    - spatial modification (pan etc.)
    - spectral modification
    - level of individual sources
    - effects (reverb / delay etc.)
- graphics overlay
    - static / animated graphics
    - driven by data
- concurrent coverage on same screen
    - Picture in Picture
    - Split screen

Certain aspects of AAF and IMF may be good starting points for such a data model, with the AAF Edit Protocol and IMF Composition Playlist having particular relevance here. The EBU Class Conceptual Data Model (CCDM) and EBUCore metadata set also have a part to play. There were also suggestions from the group that the capabilities model of SMPTE ST 2071 (Media Device Control) may have something to offer in this space. It was noted that MPEG DASH manifests can be used to achieve a certain level of interactivity or real-time adaption, and that these manifests (at least partially) map to IMF composition playlists. However, we should be

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

wary of being constrained by the limitations of delivery mechanisms or transports in the definition of a common interoperable model.

Taking a more philosophical perspective, it is not always clear whether compositional metadata should be regarded as control plane data or whether there are situations in which they should be treated as essence (i.e. on the data plane). At the point of generation, it seems clear that control events produced by a user interface are on the control plane, but when this data is sent downstream as part of a compositional description alongside the media flows it could be argued that it is being treated more like data-essence. This distinction may not be particularly important unless the requirements for transport of control and data in a given system are in conflict (or are assumed to be different).

Some doubt was expressed about whether we could hope to capture all necessary contextual information to fully inform downstream automated production operations, which leads to the question “what is the minimum amount of context required?” It was suggested that ongoing work in SMPTE on Open Binding of IDs to media (OBID) may provide a “line in the sand” in the quest to answer this question. A brief résumé of the scope of this work was given:

- Focussed on tracking and monetising commercial messages
- Open, Standard means of embedding identifiers e.g. for ad insertion
- Every show has an ID, every ad has an ID

In the context of our WebFirst scenario, the following comments were made with respect to SMPTE OBID concepts:

- Use cases encompass not only traditional ads but virtual billboards, product placement, sponsorship messages.
- In a live context, can only be managed by capturing production decisions rather than explicit metadata.
- Production decisions used in conjunction with other data to infer presence of advertising.

Multi-platform, responsive and personalised experiences

Ultimately, the EUX may be delivered to the end user’s device as a set of components, with final rendering occurring in the browser, tablet or set-top box, enabling a high degree of user control over the experience. In principle this defers all the rendering of the experience until the last possible moment, allowing the compositional metadata to be tweaked directly by the user. In practice it’s more likely that a highly constrained set of options would be offered to the viewer as part of a curated experience built from pre-produced elements. There are a few main reasons for this expectation:

1. Unnecessary complexity impacts usability for the viewer (how many in the audience really want the bother of producing their own programme?)
2. It is likely that a broadcaster would want to retain some level of editorial control over the end result, while offering extra choice and flexibility to the viewer.
3. Streaming all possible source elements would be highly inefficient and almost certainly prohibitively expensive to distribute.

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

The benefit of this approach is that, within predefined limits, the user can optimise their experience to their viewing / listening environment and/or personalise it according to taste. Personalisation may include choice of camera angle or option to replay from a different angle, or ability to modify the sound mix, for example to change the balance of commentator and ambient sound, or bias the crowd noise towards the viewer's team's end of the ground in a football match. Audio personalisation in particular may be of immediate practical value to those with hearing difficulties, or when listening in a noisy environment.

One of the advantages of 'The Internet' as a delivery mechanism is that it's bidirectional, providing us with the opportunity to pass information back up the chain. The back channel is already used in an oblique way by internet broadcasters to gather detailed metrics from the audience about viewing habits. This information may be used for ad targeting, or to inform future commissioning decisions. It is common in text-based web publishing to use live A-B testing, delivering several different versions of a page, headline or clickable link to a section of the audience, using analytics to determine the most popular version and converging on the favourite. This technique is already being applied to video advertising on the web in limited ways. How long will it be until we see it applied to internet-delivered TV, like real-time movie focus groups?

The experience delivered may also be tailored to the device on which it is being consumed. For example, less capable devices may be served a variant with fewer personalisation options; devices with smaller screens may use more close-up shots or overlay the graphic elements in a different way. This form of adaptation is similar in concept to responsive web design, where web pages are designed to dynamically change their layout for different sized screens or browser windows. To do this the distribution service needs to be able to identify (or at least infer) the client device type. Reporting of client type is standard practice for web-based distribution, as it's built into the core protocols of the web.

As yet there are few examples of the back channel being used in real time to influence the timeline or structure of a show beyond simple voting applications or reporting of social media (e.g. reading out tweets on air), and this is generally achieved using separate websites or apps running on companion devices. As internet-based distribution becomes more prevalent there will be significant scope for more creative, integrated use of this facility to gather information from the audience, allowing live production teams to respond in real time. One example brought up during the group's discussions of a show with a relatively open-ended / loose structure that is altered during recording is the Dr Phil TV Show. This show is recorded before transmission, but it is easy to see how a similarly fluid approach to production could make real time use of aggregate input from a live TV audience.

Social Media, Artificial Intelligence and Machine Learning

While it is unlikely that a single viewer will spend the amount of time required to customize every viewing experience, several media companies contacted during the compilation of this scenario reported that they have experimented with creating and delivering customized content to each viewer, or to blocks of viewers based on using aggregation and analysis of data available from a variety of sources. These techniques are being used in anticipation of a time when:

Live analysis of social media allows viewer sentiment to influence the automated production of different versions of a live event (e.g. different camera angels, insertion of different pre-produced

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

pieces during breaks in the action, use of different audio elements and announcers) for groups of viewers or even a single viewer

Aggregation of these data with additional information that has been previously learned about the viewer's preferences, or which is based on previously learned viewing behaviours in particular regions (e.g. preference for a particular football team, or a locally favoured player) that may be used to automatically produce customized viewing experiences.

A viewer creates a live remix of an event for his friends down at the pub (or for hundreds of thousands or millions of "friends", as has been done countless times with file-based content on YouTube)

Key trends which are driving these possibilities, which are already being leveraged for Internet applications include:

Big data integration leveraging location, social and situational awareness

Application of the concept of the file-based package (a wrapper containing a number of different pieces of video, audio and data content, along with rendering instructions) to live scenarios, where multiple content elements are distributed to different devices, along with render instructions which are executed at the end-user device

Offering the end viewer options regarding the viewing experience based on that viewer's past viewing preferences

Leveraging best practices in Computer Science around proper layering in system design, virtualization, atomic functionality, identity, composability and reusability to create cloud-based systems that can quickly scale up or down in functionality in real time (or even ahead of real time in anticipation of heavier loads)

Human Factors

WebFirst production presents a number of technical challenges, but we should appreciate that this is at most only half the story. Live production teams are generally tight-knit units that have developed efficient ways of working together to produce a particular type of content. Many of the issues that must to be addressed to satisfy the requirements of our WebFirst scenario are human factors that have little to do with technology. Worse, introducing radical technology-driven workflow changes can be highly disruptive to the smooth running of an experienced team, making technology part of the problem rather than the solution.

The group spent some time discussing this topic, attempting to identify the most important human factors and exploring how these might be addressed, through the use of technology or otherwise.

One of the biggest questions for WebFirst production is how to deal with simultaneous production of multiple versions whilst maintaining consistently high production values for all platforms - a particularly acute issue for live production. The most extreme solution might be to replicate the entire production team for each version required. This is clearly impractical on various grounds (cost and space, to name but two), but leaves the question of how a production team *should* be structured to satisfy the demands of WebFirst. Can the use of specialised

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

algorithms for semi- or fully-automated versions out of a basic version help? Of course the answers will be dependent on the characteristics of the production and the design of the experiences that are built from it. How different are the experiences targeted to each platform? How do the differences in experience design and audience expectations of each platform affect the source elements that go into these experiences? How much personalisation is offered in each case? Does the content need to be delivered to every platform live or to some only close-to-live, as packages or clips?

The group tried to navigate this minefield by identifying requirements that are likely to be common across a wide range of use cases within the scope of our scenario, as well as considering some existing products targeted at multi-platform production.

Imposing some Structure

The more the structure of the production can be tied down ahead of time the less there is to do in the production gallery on the day. This is true regardless of the number of platforms targeted, but planning becomes more critical in the multi-platform case. Understanding how the requirements of each platform differ and how content is expected to be reused across the platforms is crucial. Platforms requiring a markedly different style of presentation or relationship to the live timeline may need a separate producer.

The Role of Metadata

Machine-interpretable production metadata can aid decision making, whether the decisions are made by machines or by humans (since machine-readable metadata can easily be converted to visual indicators on production user interfaces). The potential scope for discussion of the applications of metadata is huge, so we’ll have to make do with a few simple examples.

1. *Calling the shots:* In a live production with multiple outputs the director no longer has complete control over which camera feeds are visible to the viewer at any one time. This may subtly change the responsibilities of the director, but it is also more important to have good information about the ‘validity’ of each feed. Signalling from camera operators to communicate when they’re in the process of reframing could be combined with metadata generated from video analysis designed to assess image stability, focus and exposure to provide an indication to downstream systems of the usability of the feed, without relying on the director looking at the pictures on a screen. Inversely, signalling from automated ‘sub-directors’ could instruct camera operators which shots to hold, e.g. until the sub-director has switched to a different camera.

2. *Who’s up?* Descriptive time-based metadata identifying the event, participants, presenters etc. enables indexing of stored media for segmentation purposes. This metadata might be generated from the combination of schedule data, video and audio analysis, and perhaps augmented or fine-tuned by hand. Making this metadata available downstream alongside the content may aid in content navigation in the EUX.

3. *Virtual billboards:* Billboards around the pitch are tracked in the video and the affected screen coordinate sets tagged with IDs. These will be overlaid with virtual billboard adverts targeted according to information gleaned from web browsing history and viewing habits. The end user device ties up the region tags with IDs on the pre-delivered ad

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

content and performs the rendering of the virtual billboards into the video using its 3D graphics processing capabilities.

It's essential in all of these cases that the metadata is machine-readable, as this facilitates automated processing. These examples also highlight that the metadata may be used anywhere in the chain, from production gallery to end user device.

Lightening the Load

A theme that emerged strongly in our discussions was the need for automation. As productions get more complex, with more diverse and flexible delivery requirements, automation becomes crucial as an aid in producing consistent quality across all outputs through guaranteed repeatability with minimal intervention.

As the astute will already have understood, automation is closely linked with metadata of various types, and here lies a potential conflict; generation of metadata often represents an additional overhead on members of the production team. While this is to a certain extent inevitable, and some mind-set adaptation may be necessary on the part of production professionals to reassess the value of metadata, it is incumbent on production system designers to minimise the overhead of metadata generation at every stage. We should always be mindful that technology should speed things up, not slow things down. In many cases useful metadata can be generated through analysis of the media itself, or by inference from data captured from several sources in ways that impose little or no extra overhead on production staff (e.g. existing production interfaces, sensors, single button interfaces). This not only frees up the production team to focus on their primary responsibilities, but also works around the issue that manual entry of metadata can be error-prone, particularly in high-pressure situations. Metadata frameworks and templates can also be a useful organisational principle; these rely to some extent on skilled domain-specific data modelling to identify the elements that can be prepared in advance for a particular production to minimise the overhead in the heat of the live event.

Some production professionals will be resistant to the introduction of automation, seeing it as a threat, either that their role will be usurped by a robot, or that their creative vision will be compromised by an algorithm. The challenges of WebFirst production will undoubtedly change the traditional responsibilities of production roles, and some of that change will be as a result of the increased use of automation, but it is not intended (at least in this context) to fully replace the roles of skilled production personnel; rather to provide assistance in making production complexity tractable.

Tools for Multi-platform

The group considered a number of existing tools designed for the production of web-delivered or multi-platform content. There are undoubtedly plenty that were overlooked.

- never.no
  - multi-platform production environment
  - automated tools to assist production personnel
  - leverage automated metadata where possible
  - reduce overhead while allowing creatives still to be in control
- watchwith
  - adding time based metadata to video

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

    - identity of live streams helps with this in moving it closer to real time
    - reliant on asset being stored so it can be identified
    - limiting factor may be how fast metadata can be added
    - not real-time, can be near real time
    - use within a multi-platform distribution environment
    - relies on personnel to add metadata
    - target advertising more effectively
    - feedback loop from audience or within a distributed organisation
    - real time?
- telestream "tag & bag" tools
    - metadata schema to support multi-platform content creation
    - using metadata to drive workflow
- Various providers of “white label” video player apps, re-skinned by content distributors, avoiding the need for content distributors to develop apps of their own (and in many cases to manage internet-based distribution).

Regulatory and Commercial Landscape

The group raised a number of issues related to WebFirst production in a regulatory, legal and commercial environment that has been evolved for linear broadcast. In common with the wider World Wide Web, some of the features of WebFirst video production and distribution are difficult to reconcile with prevailing rules and regulations. Some of the issues may be mitigated through the application of technology, but in many cases it is new technological capabilities themselves that are at the heart of the matter.

- Regulatory compliance (e.g. KidVid FCC regulations, epilepsy, editorial bias (for PSBs)). How are these measured when everyone’s experience is different?
- Increasingly complex commercial landscape where content is originated, delivered by different companies (with platforms often owned by other, different companies), in contrast to the current situation where many broadcasters are in control of the whole chain from origination to end user device.
- Geo-specific content rights are already causing issues for streaming services such as Netflix. Delivering flexible experiences that may be composed from content streams derived from different sources, with different rights (or different rights models), where the end user has some choice about which components are included, make for an even murkier situation.

Conclusions

Elemental Essence

WebFirst production requires simultaneous, independent streaming of audio, video and data elements in multiple encodings and resolutions. This preserves the necessary level of flexibility to dynamically construct End User eXperiences (EUXs) based on different combinations of available media at any point in the production and distribution chain. The choice of protocol used for transport of the essence is dependent on a number of factors, including latency constraints for different use cases, reliability requirements and suitability for delivery to different technology platforms (e.g. web browsers don’t deal with RTP directly). In some parts of the system the adaptive streaming capabilities of protocols such as MPEG-DASH may be beneficial. Timing and

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

identity of streams should be independent of essence type, format or transport protocol so they can be synchronised wherever necessary using a universal approach.

Identity

Much of what has been discussed in this document requires a strong content identity model to enable tracking of essence and data through the production process and to facilitate the dynamic construction of End User eXperiences from those elements. Identity also provides the foundations for navigation of the EUX for the consumer, as well as providing a solid basis for semantic tagging so that external content can be associated by conceptual similarity.

Timing

To support seamless transition from live to playback from storage, absolute timing against a common real time reference clock must be recorded and bound to the media to be used for live cross-media synchronisation, and indexing once it's in the store. The reference clock must be of sufficient resolution and accuracy to support simultaneous video capture at different frame rates (e.g. HFR for slo-mo action replay alongside regular frame rates). Precision Time Protocol (IEEE1588) provides ample resolution and accuracy for this purpose. The live media transport infrastructure must support concurrent carriage of media of different types, formats and rates. Indexing of media in the store must use a consistent addressing mechanism across all flows, regardless of type, format or rate.

Data

We can't hope to capture all possible future requirements for data representations. Or current requirements for that matter; there are too many variations and the landscape is likely to change very quickly. This is one area where a rigid, centrally-standardised schema is probably not the answer. It would be much better to define a generic, minimal schema that can be extended as necessary.

A framework that enables free association and aggregation of time-related data feeds, allowing derived feeds to be used as control inputs to a video, audio or data processing device, would be an invaluable tool for production automation. One way of achieving this would be for data and control events to share a common container format, allowing data feeds sourced from direct measurement or media analysis to be used directly or indirectly as device control inputs.

Interoperability

Fully-fledged WebFirst production requires that different combinations of elements of an experience can be performed anywhere between capture and the end user device, with metadata communicating how the elements should be composed into an experience passed downstream alongside the media. The rather onerous-seeming requirement that follows from this is that tools and devices throughout the production and distribution chain must speak a common 'language', providing interoperability on both control and data planes. Before this is pronounced as being unachievable, it should be pointed out that this level of interoperability is already expected from (and achieved by, to a large extent) web browsers delivering dynamically-rendered text and graphics-based experiences.

This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported

Human Factors

The most important message to take away from our analysis of human factors for WebFirst production is the need to counteract complexity as the demands on the production team multiply. This may be achieved by the application of technology, but that technology should be as 'invisible' as possible. The last thing that production teams need is for technology to force a workflow on them that involves extra responsibilities during a live event. In the design of any media production system, streamlining the user experience should be paramount.

What is JT-NM?

The Joint Task Force on Networked Media (JT-NM) was formed by the European Broadcasting Union, the Society of Motion Picture and Television Engineers, the Video Services Forum and the Advanced Media Workflow Association (AMWA) in the context of the transition from purpose-built broadcast equipment and interfaces (SDI, AES, crosspoint switcher, etc.) to IT-based packet networks (Ethernet, IP, servers, storage, cloud, etc.) that is currently taking place in the professional media industry.

The Task Force was set up to foster discussion among subject-matter experts and to drive the development of an interoperable network-based infrastructure for live media production, encompassing file-based workflows. It brings together broadcasters, manufacturers, standards bodies and trade associations.

The JT-NM counts more than 300 participants from 175 organisations. More information on the JT-NM including its scope and previously published works may be found on our website jt-nm.org.

Contributors

Sub-group chair: Robert Wadge (BBC R&D)

Participants: Christian Adell Querol (CCMA), Bryn Balcombe (London Live), Chris Connolly (NBCU), Frans de Jong (EBU), Mike Ellis (BBC), Jean-Pierre Evain (EBU), Roger Franklin (Crystal Solutions), Janet Gardner (Perspective Media Group), Brad Gilmer (Gilmer & Associates), Jürgen Grupp (SRG), William Hayes (Iowa Public Television), Sara Kudrle (Grass Valley), Nilo Mitra (Ericsson), Félix Poulin (EBU), Andy Rosen (Consultant), Kuldip Sahdra (VIXS Systems), Thomas Saner (SRG).

(cc) BY-ND This work is licensed under a Creative Commons Attribution-NoDerivs 3.0 Unported