Study Mission report on Object-Based Media

DVB Document S102

May 2025

Intellectual Property Rights

Please refer to the IPR policy of DVB Project available at: https://dvb.org/about/policies-procedures/

Foreword

DVB is an industry-led consortium of broadcasters, manufacturers, network operators, software developers, regulators and others from around the world committed to designing open, interoperable technical specifications for the global delivery of digital media and broadcast services. DVB specifications cover all aspects of digital television from transmission through interfacing, conditional access and interactivity for digital video, audio and data. DVB dominates the digital broadcasting environment with thousands of broadcast services around the world using DVB specifications. There are hundreds of manufacturers offering DVB-compliant equipment. To date, there are over 1 billion DVB receivers shipped worldwide.

Contents

1. Introduction & Overview

Inspired by trials and service launches worldwide, The DVB Project is undertaking a study mission to gain insight and collect use cases for Object-Based Media (OBM). The term object-based media is used to describe an approach to producing, distributing and consuming media content that uses separate media assets, known as ‘objects’. An object consists of essence and its associated metadata, which describes the essence.

For example, a viewer watching a football match might choose to hear commentary from a pundit who is a supporter of their team, instead of the standard broadcast commentary. By producing three audio objects, one for each team and one standard broadcast, alongside the video object, this choice can be offered to the viewer. Interaction by the user can occur explicitly, through user controls. The content may also be automatically adapted based on user preferences or device characteristics. As well as offering new opportunities to adapt the narrative experience, object based media also offers opportunities to optimise the presentation of content for the device being used or the environment in which the content is being consumed.

Where object-based media presentations are used, some objects are considered essential to the default content, such as the standard audio or video, while other objects might be additions, such as an alternative language.

The situation today

Most standard programmes today contain at least two additional media tracks (e.g., audio description, subtitles) but many contain several more (such as visual signing, director's commentary, additional graphics, EPG links, trailers and references to online information).

These can and sometimes are considered to be objects, and can sometimes be combined to offer several experiences.

Although these tracks exist today, they are not currently considered as object-based media as each one is a stand-alone service which can be enabled or disabled and has limited metadata, and they rarely have the metadata that describes how they can be combined to form an object-based presentation.

Recently however, it has become possible to create and deliver more sophisticated object based media, such as branching narratives, personalised presentations, which offer totally new viewing experiences. Many of the underlying media types are delivered in existing standardised ways, but today the methods used identify them and link them to the primary programme or to each other often rely on proprietary or app specific data.

In current DVB systems there are specifications for carrying the most common types of media essences, e.g., subtitles or alternate audio tracks, but these specifications are designed for that specific purpose and a common method to label, identify and carry metadata for multiple objects is not available.

Preparing for the future

As media and media delivery systems evolve, the number of elements associated with the content increases. Social media posts, trailers, promotional web content, photos, sound clips and a host of other assets are now associated with media. All these elements have traditionally been managed in an ad-hoc way and the media archive for many programmes in the past, merely consisted of a master copy of the programme itself, with a short filename, or label on the tape or storage media.

In addition, the media may be composed dynamically from separate digital media assets, and can be made responsive to user preferences or interaction, and rendered dynamically by the client device or by a cloud service.

The group is examining how “objects” consisting of media elements with metadata are being used in the industry and determining how well existing specifications cover the carriage of multiple objects, and whether

there are commercial drivers that justify further standardisation by DVB. This includes the carriage of the underlying media essence, its metadata, and re-assembling the experience "rendering" at each stage of delivery.

Whilst we focus on the distribution of the media, questions of ease of use, accessibility, user experience, interaction and enabling presentation to and selection by the user are also in scope.

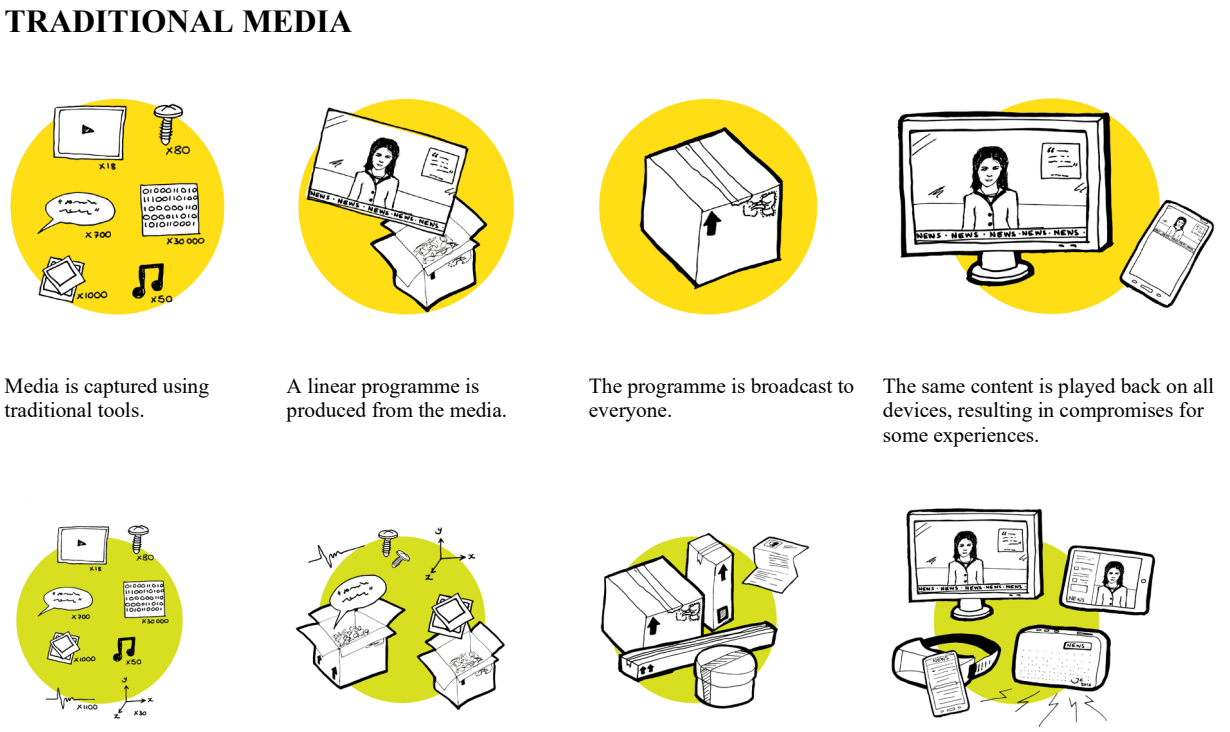


Figure 1. OBM Explained - Image courtesy of BBC R&D 2018

1.1. How can Object Based Media Be Consumed?

To present object-based media, it is necessary to decode the essence using instructions provided by metadata delivered alongside the essence. "Rendering" is the term used to describe this process of assembling all the relevant objects, using the constraints imposed by the metadata, the playback environment, and the user (either directly or via their preferences), to create the final media experience to be consumed by the user. Rendering is required throughout the value chain, during production, playout and consumption of the media.

1.2. Presentations to the group

The following industry experts presented their work to the group:

- Ru Howe – Stornaway.io – FAST and OTT use cases
- Dave Marston - ADM
- Andrew Dunne – IMF User Group
- Mathew Brooks – BBC media player for the World Cup
- Adrian Murtaza – Fraunhofer OBA Brazil
- Marcin Mazurkiewicz - Dolby OBA Poland
- Iraj Sodagar -MPEG presentation on the new DASH random access support
- Ben Shirley - Salsa Sound / Salford University - Sports use cases for OBM
- Chris Needham - W3C/BBC OBM support in W3C specifications
- Rajiv Ramdhany & Nicholas Race – AI4ME – overview of work programme
- Xu Yiling - Shanghai Jiao Tong University – OBM and SMT
- Ruud van der Linden - Co-Founder - Infuse Video

2. Glossary of Basic Terms

For the purposes of the present document, the following terms apply:

The definitions hereunder are to be read in the context of this report but should broadly correlate with those already used by DVB.

Object-based media:	Object-based media describes any approach to producing, distributing and/or consuming media content (Essence) which uses digital media assets, and metadata which explains what they are and how to use them, These are known as 'objects'. Objects may take one of three forms: 'layers', 'chunks' or 'transmedia'. **Layers**: Layers are either designed to be consumed together with or instead of other objects at a moment in the content timeline. Layers may be combined in groups to create alternative or augmented experiences. They must be co-timed and synchronised to be meaningful. Examples include layers to carry subtitles, graphics (e.g., for sports programmes), signing, audio description, additional commentary or languages, and effects such as separated sound, for example to change dialogue loudness. **Chunks** or **Object Media Clips (or media segments, components, fragments)** : Chunks are a sequential segment of a programme which can be manipulated by consumers and/or producers to create a more flexible viewing experience, for example to create programmes of differing lengths, with branching narratives, or for more personalised content or advertising - pre-rolls, mid-rolls, post rolls, intros, credits, bumpers, tops and tails are also examples of chunks **Transmedia Objects** (ancillary deliverables/assets): Transmedia objects are supplemental to the content itself and are not directly part of the consumption process. They can be consumed separately to add to the overall experience. Examples of these include objects containing trailers, 'making-of' documentaries, outtakes, cast and crew interviews, location information, reviews, scripts, social media pages, blog posts etc. They can even extend to objects containing or linking to a related radio drama version, merchandise, e-magazines and spin-offs. All the underlying media elements exist today but are very rarely arranged, ordered, tagged, labelled and searched as a true object database,
Metadata	Metadata is data that describes other data. In the context of OBM, media elements alone hold no purpose unless they are accompanied with metadata to explain if, what, where and how they can be used.
Rendering	Rendering is the process of assembling all the relevant objects, using the constraints imposed by the metadata, the playback environment, and the user (either directly or via their preferences), to create the final media experience to be consumed by the user. Rendering may be done either:

	i. during the production process (as for a traditional broadcast); or ii. in the cloud or at the network edge; or iii. directly by the consumer's device. In cases (ii) and (iii), the only objects that need to be delivered are those required to create the media for the relevant consumer.
Companion device	A separate device from the main (TV) viewing screen, connected via IP, such as a phone, tablet, laptop or bespoke accessibility hardware (e.g., a single switch interface), that is used to augment the main viewing experience, for example by having some objects made available solely on that device. This would offer some users an enhanced experience, e.g., for sports match statistics, without disturbing the viewing experience for other users of the main screen.
Essence	(Sometimes also known as Content) consists of elements of a program which is presented to the consumer (Audio/Video/Graphics/Haptics/EPG data/Audio Description/Subtitle text). Metadata should not be considered as Essence.
Next Generation Audio (NGA)	Next Generation Audio (NGA) Systems provide Broadcasters, Operators and Content Providers with more flexibility to create and deliver their content. These NGA Systems introduce a number of new concepts and techniques, including: ● The capability to provide Immersive Audio with the addition of height elements. ● The capability to provide personalization options to broadcasters and consumers.
Targeted Advertising	The substitution of an advert from a linear broadcast with an advert chosen by an advert decisioning service based on criteria such as viewer profile, viewing behaviour, environmental or contextual factors.

2.1. Abbreviations

For the purposes of the present document, the following abbreviations apply:

DVB CM	DVB Commercial Module
DVB TM	DVB Technical Module

3. Terms of reference

These terms of reference were agreed by DVB as the basis of the work of the group:

"To review and analyse how programmes containing multiple media elements can be carried in DVB systems with a focus on IP delivery but not excluding TS delivery and identify any potential need for new or revised specifications to support Object Based Media. This will include the carriage of media objects (with associated metadata), and the support for rendering at each stage of delivery. The work will be driven by commercial use-cases, and will consider what is feasible today, and any shortfall. The study mission will examine whether or not the industry is on a roadmap towards an OBM approach to distribution, and judge if and when any DVB output would be needed to align with this. Whilst the Study mission will focus on the distribution of the media, questions of ease of use, accessibility, user experience, interaction and enabling presentation to and selection by the user will also be in scope."

4. Current state of OBM in production and distribution

4.1. Major operators and broadcasters

Presentations to the group describing trials by BBC, BT, Sky and others, demonstrate trials of aspects of OBM such as use of Next Generation Audio (NGA) to allow personalisation of the dialogue level. Additionally, SBS, TVP and TVGlobo now have commercial services using NGA. Although the term OBM is not always used, the use of multiple elements, personalization, metadata, and the ability to render different versions of the content from its elements are present and therefore represent a full or partial OBM workflow.

It is also clear from the survey that most respondents including content creators and distributors consider OBM important to the future of media delivery.

4.2. Major Streaming platforms

Services from Peacock, Netflix, Amazon, YouTube and BBC iPlayer amongst many others already support many OBM elements, albeit currently in proprietary ways in their own applications. Many of these services support both layered objects, and branching narratives, and some now support a wide array of transmedia with rich metadata about the content too (e.g. Amazon X-ray).

4.3. Browsers

The web already supports some aspects of object based media. HTML enables web pages to be composed of text, images, video and audio. Cascading Style Sheets (CSS) enables responsive positioning and layout of elements in a page (e.g., to adapt to different screen resolutions), as well as animations and transition effects. The JavaScript and WebAssembly languages provide programmability. The HTML Canvas API allows programmatic 2D bitmap and 3D graphics rendering, and web pages can include Scalable Vector Graphics (SVG), an XML based markup language for vector graphics. These technologies allow developers to create web applications across different device form factors and operating systems.

Web technologies are widely used in TV devices, e.g., through inclusion in TV specific standards such as HbbTV. The Web Media API Snapshot (a joint publication of the W3C and CTA) lists the web APIs to support media web applications that are supported across all of the most widely used user agent code bases. The intent of the Snapshot is to provide a documented baseline set of specifications for device manufacturers that include an HTML user agent (including televisions, game machines, set-top boxes, mobile devices, and personal computers) to develop products that support the APIs. The Snapshot is updated annually.

In practice, the set of features included in TV based browser engines typically lags behind desktop and mobile browsers, sometimes by a year or more. Desktop and mobile browsers are updated regularly, e.g., on a 6 or 8 week cycle, whereas TV browsers tend not to be updated to add new features once a device is released. Manufacturers typically include an updated browser engine as part of each new product release.

Refer to section 6.5 for more detail on relevant web browser standards.

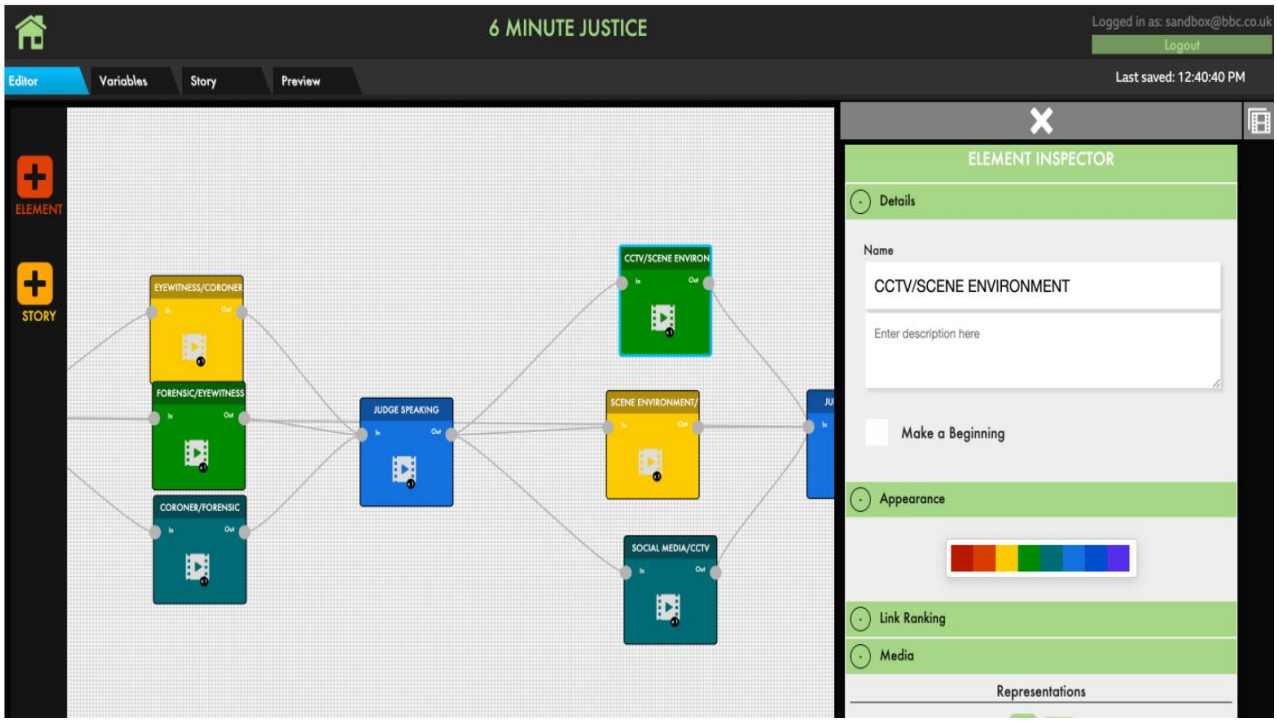
4.4. Narrative Tools

Recently, a number of production tools have been developed both by industry and by academic bodies that are designed to simplify the burden of asset creation for OBM and embed practical design principles into the process of OBM content creation. Many of the tools are not commercialised but adhere to principles of open-source distribution and use. This indicates a recognition within industry of a need to consolidate toolchains

and create working standards that allow broadcasters and independent production companies and creative organisations to speak the same language and work to the same standards. Many broadcasters we have spoken to have indicated that they would rather be in a position to use and/or adapt existing industry standard tools but are generally forced to create solutions from scratch in order to explore all aspects of OBM for content. We have also been made aware of many proprietary solutions related to OBM being developed by broadcasters across the board, particularly in areas of AI metadata recovery and editorial decision-making support. The following is a list of publicly or commercially available software supporting OBM adoption. BBC R&D has been particularly active in examining how chunks could be used to create new content types and audience affordances, as part of its public broadcasting remit.

4.4.1. StoryFormer production tool

In order to provide better tools for content producers within and outside the BBC it has created a tool called StoryFormer.



Screenshot from Storyformer tool which shows how layers and interactive narrative chunks are visually organised. (Source: BBC R&D, 2019)

StoryFormer is a BBC R&D tool for creating flexible, responsive stories. It uses branching narratives in its editing user interface, to allow producers of content to create complex relationships between audio and video clips without advanced technical knowledge. The tool and its makers offer an expanded idea of interactive narrative that includes:

- Standard 'lean-forward' interactive features in programmes
- Lean-back viewing where environment, default preferences, light-touch decision-making or other variables inform media playback.
- Choice of media type: giving the audience a choice of playback as audio only, video or graphic novel.

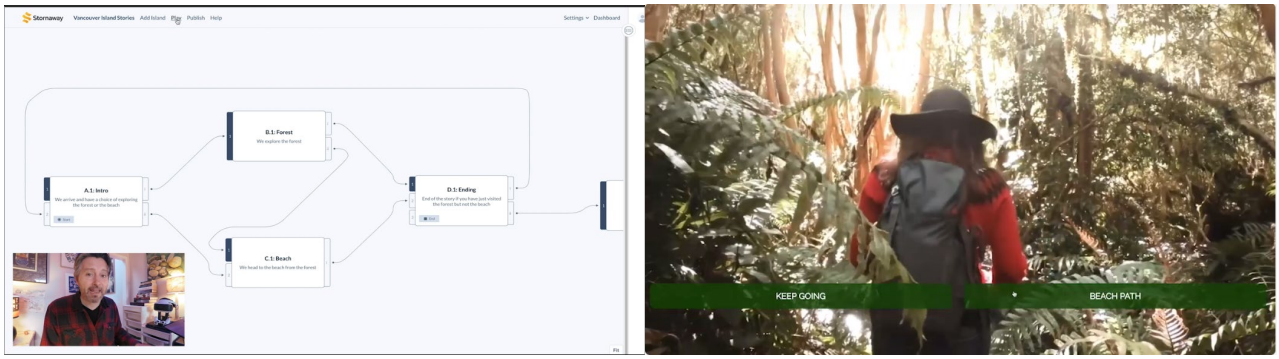
Storyformer was used to create an interactive version of technology programme Click which imagined a future where "personalisation and choice play [...] a bigger role in how we watch media" (Beckett,

2019). In it, users could choose the topics they explored more and those that they were less interested in, creating a programme that was personalised to their interests. According to a BBC case study on the production, approximately three hours of video could be watched in over *84 trillion ways*. In terms of delivery options for producers within the BBC, whilst it is not yet possible to publish OBM media from Storyformer to the broadcaster's iPlayer or other hybrid offerings it recently became possible to publish this content on any BBC webpage as an interactive video. More work is going into developing the editing capability of Storyformer going forward.

Storyformer is part of the MakerBox toolkit developed by the BBC under a closed beta licence, individuals and companies wanting to trial the tools must register with the BBC and agree to use the tool only for non-commercial purposes without agreement. The underlying data model and core player are open source (GPL-3). Other tools using object-based production methods within MakerBox include the BBC's Audio Orchestrator which allows users to connect multiple devices - smartphones, tablets and laptops - to play out a fully immersive audio experience and Charisma.ai's platform that allows the creation of voice-controlled storytelling powered by AI (https://charisma.ai).

4.4.2. Stornaway.io

UK start-up Stornaway are also creating tools and workflows that allow content creators to quickly create and distribute stories that take advantage of the existing branching narrative capabilities of YouTube, browsers, and games engines such as Unity. Their tools are also compatible with the open source tools provided by the BBC.



Stornaway.io - Interactive video software | Creative Storytelling Tools for interactive video stories and games | Stornaway.io

This video clip shows how a branching narrative video can be easily created

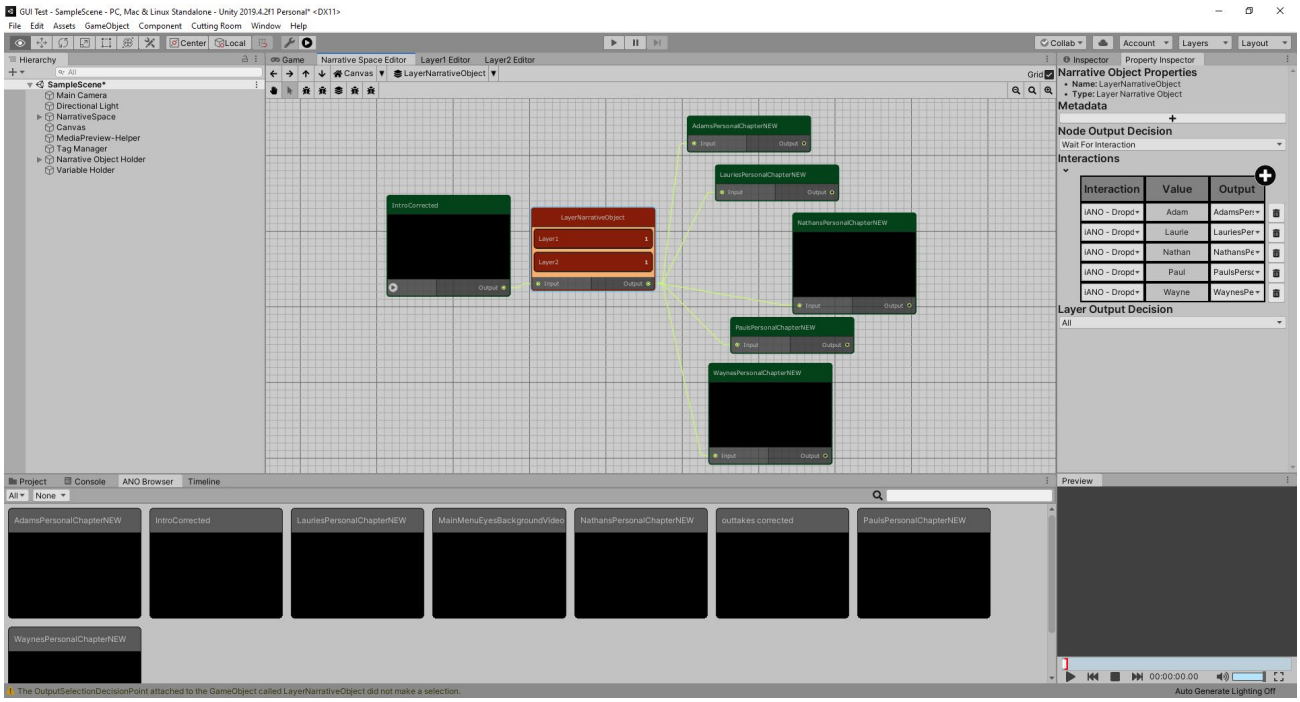
https://www.stornaway.io/wp-content/uploads/2021/01/Stornaway-site-Story-map-clip-v2-720-2.mp

They report a very strong and growing level of interest in creating simple stories that can be used in education and commerce as well as entertainment.

As with most branching narrative playback the web browser or app are the most capable for this type of media with less interest and more complex integrations needed for TV or hybrid platforms.

4.4.3. Cutting Room

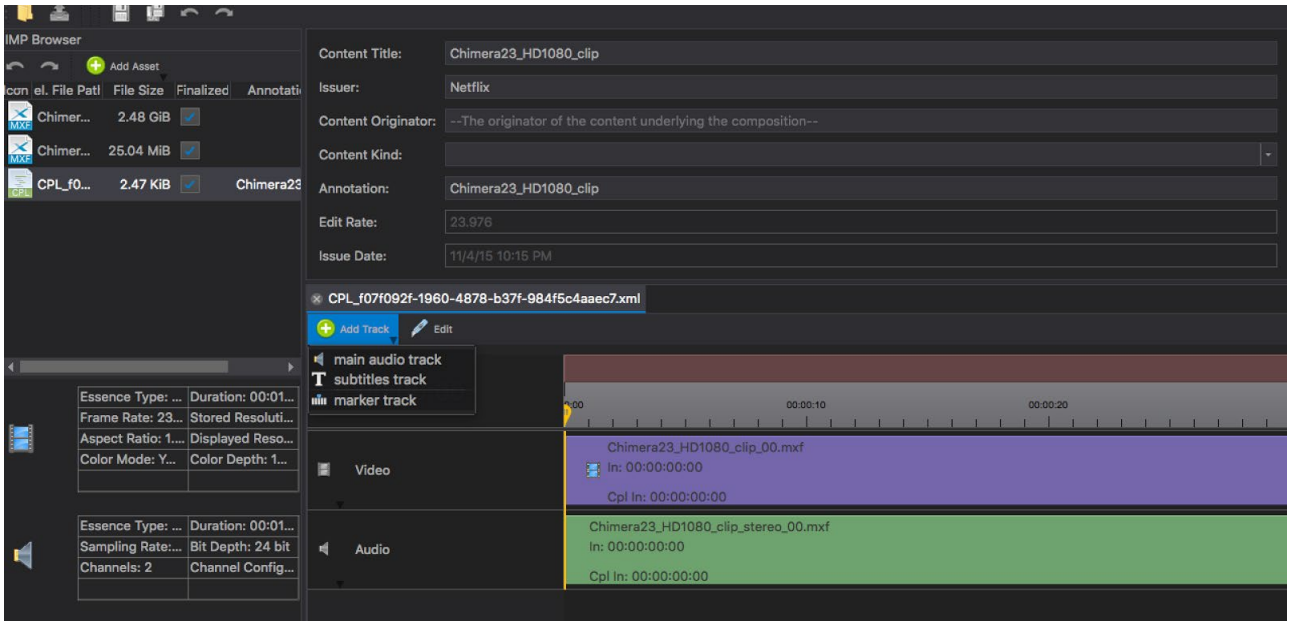
A similar tool by the University of York called Cutting Room was developed in partnership with BBC R&D and funded by the EPSRC. A more advanced version is now being deployed in several University of York XR Stories projects (https://xrstories.co.uk), including an interactive podcast and an interactive video narration for the National Media Museum in Bradford. Cutting Room allows for the creation of both 'atomic narrative objects' that can be laid in chunks in a branching narrative sequence and/or layered on top of one another to create a 'layered narrative object'.



Narrative Space Editor with multiple Narrative Objects (NOs), including a Layered Narrative Object (LNO in orange) (source: University of York)

4.4.4. IMF CPL Editor (Wolfgang Ruppel)

The IMF CPL Editor is a cross platform tool that allows editors to open an IMF composition playlist (CPL) in a timeline and add, edit and delete audio, subtitle and metadata assets from the timeline. The edits can then be saved back to the same IMF Package. (For more information on IMF see Section 6.1, SMPTE and IMF-UG.)

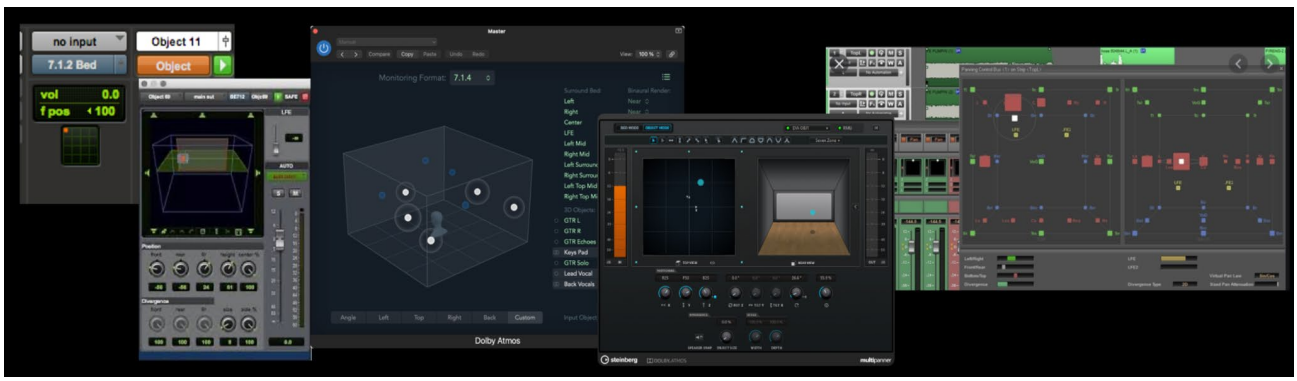


The tool is open source and development sponsored by Netflix and supported by the Academy of Motion Picture Arts and Sciences, Sony Pictures, Warner Bros., Universal Studios and 20th Century Fox (Ruppel, 2020). It is part of a wider open source tool chain supported by Netflix that includes the IMF Conversion Utility, which creates 'flat' video files such as DPP compliant AS-11.

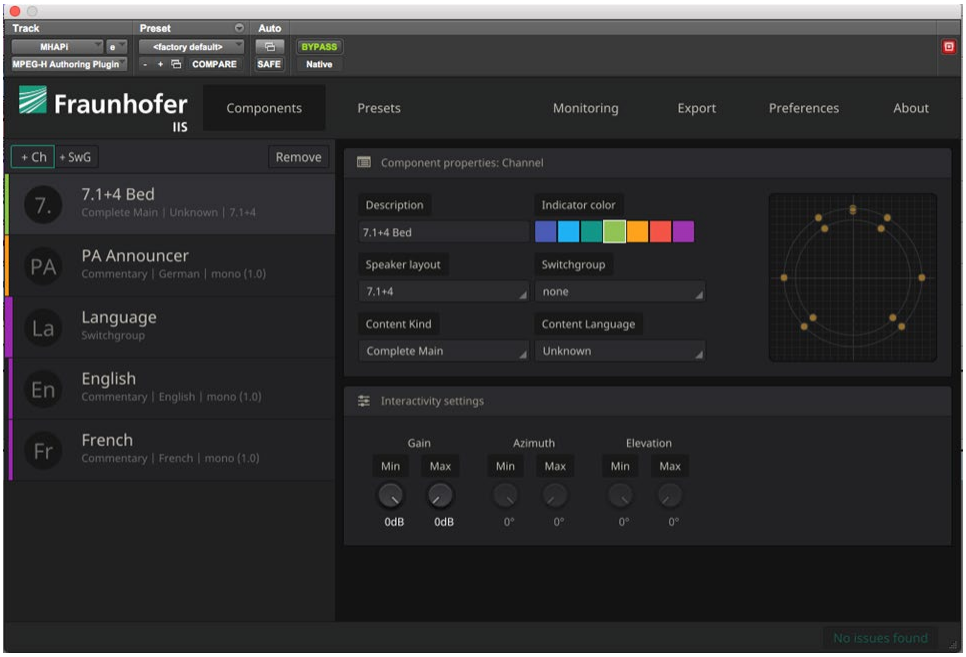
4.5. Audio Tools

Whilst immersive audio using additional height channels is now quite commonplace for some genres, especially sport, personalised object-based audio has seen slower adoption. Next Generation Audio uses objects to deliver a range of experiences including enhanced accessibility (boosted dialogue, audio description etc.) as well as a variety of creative options for the consumer to select. Regular broadcasts with audio objects using MPEG-H Audio started in South Korea in 2017 and personalised object based audio is on the air in Brazil, with successful trials in many other territories. Immersive and personalised object-based audio services using Dolby AC-4 are provided in Poland via DVB-T2, for example for FIFA European Cup '20 and World Cup '22.

The toolset is well developed, with hardware and software available from several suppliers to support live and post-produced content. Fraunhofer provides a range of free software tools for authoring MPEG-H Audio; plug-ins for Digital Audio Workstations (DAWs) and stand-alone software for authoring, checking and rendering MPEG-H are available from them. MPEG-H Audio and Dolby Atmos are natively integrated into the common audio workstations such as AVID ProTools, Steinberg Nuendo, Blackmagic DaVinci Resolve and others.



Native Dolby Atmos integration in Pro Tools, Logic, Cubase, Nuendo, Pyramix



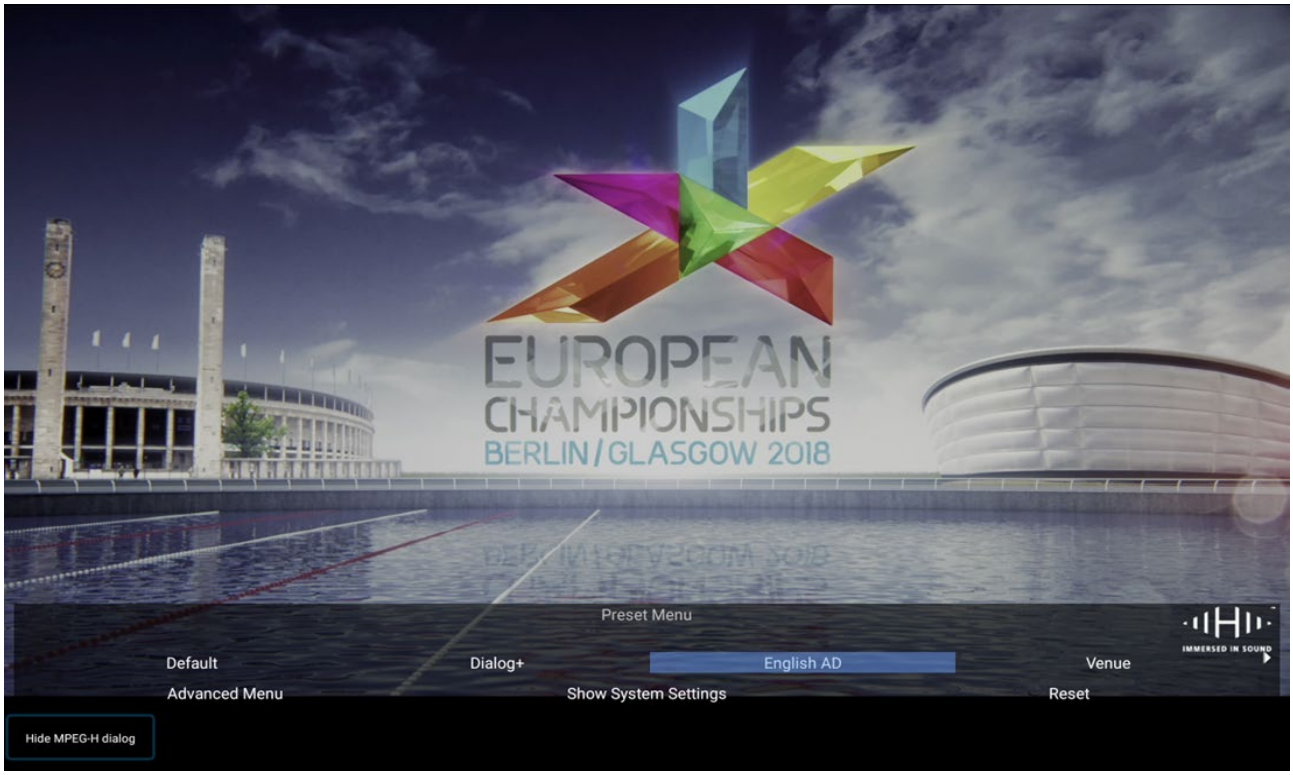
An MPEG-H Authoring Plugin

In addition or alternatively to the native workflows, plug-ins from Dolby and Fraunhofer can be used with a wide variety of digital audio workstations.

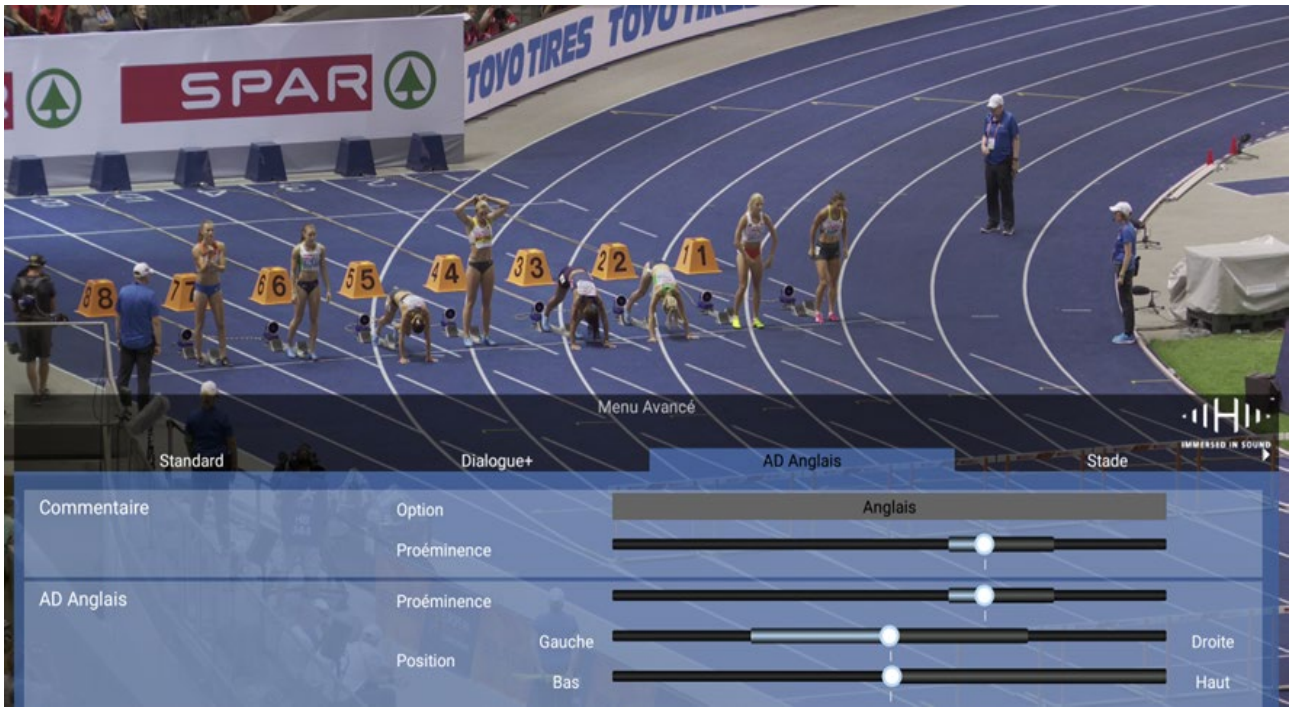
Dolby Atmos Renderer v5.0 software

Audio object encoding can be done with several solutions, including graphical user interface based, command-line based and cloud-based tools, e.g., Amazon Web Services (AWS). For live production, if hardware tools are preferred, these are available from Ateme, Harmonic, Junger, Linear Acoustic and New Audio Technology. Some live productions also rely on cloud-based live-encoding and distribution. Here, Dolby Atmos live object-based audio encoding is supported in the toolsets of AWS, Ateme and Bitmovin. Distribution coders from a large number of manufacturers also provide support for audio objects on broadcast and IP platforms. Both Dolby Atmos and MPEG-H Audio production is supported over traditional Serial Digital Interface (SDI) infrastructure and using SMPTE ST2110 standard, although the implementation of this standard is still evolving.

Hardware object based audio encoders with software user interfaces are available



An example of a user interface offering a number of pre-set options. These can change during the programme if required.



More comprehensive user interfaces are possible. The user can select the level of complexity they see.

In addition to use on TVs, object based audio can be delivered through mobile devices. A reasonably modern smartphone can decode MPEG-H audio in software and provide the full range of user controls, as well as rendering the binaural sound for use on headphones or earbuds. Dolby AC-4 processing is supported on many Android devices using a decoder implemented on the SOC. Object based audio is also used in smart speakers from Amazon and Sony and is the underlying technology behind many high-quality music streaming services including Amazon Music, Deezer and Tidal.

Some broadcasters have expressed a preference to create and archive content using the open "Audio Definition Model" (ADM) standard. This is a file-based standard for post-produced content which can be converted to a distribution format (AC-4 or MPEG-H) at the point of broadcast. ADM is the format of choice for delivering content mixed in Dolby Atmos to streaming services such as Netflix, Disney+ and Sky, and is well established in this area. The standard for Serial ADM (S-ADM) to support live broadcasting is less well developed and implemented, but trials are ongoing with strong support from the BBC and EBU. ADM is a large standard supporting a lot of options and this has hampered implementation but work is ongoing to define reduced profiles of this standard for broadcast production and to feed distribution encoders.

4.6. Adaptive rendering and edge computing

The 2021 Ofcom report on Object Based Media (https://www.ofcom.org.uk/__data/assets/pdf_file/0026/223973/Object-based-media-report.pdf) discusses the topic of rendering and edge computing at some length, it concludes:

> "*Edge computing provides the prospect of greater efficacy in the distribution of object-based media, allowing assets to be compiled at the 'edge' of delivery to the consumer, effectively providing the essential additional capacity required to deliver OBM at scale, especially around live events, such as news and sport. The closer to the consumer and the edge of the network the objects can be combined, the more personalisation and interaction is possible. Personalized and Targeted Advertising is one of the first applications of this from a content point of view*".

The 5G rollout which it also discusses is now in full swing and is improving the penetration of broadband across the UK. For the first time the traditional DTT/satellite broadcast path to people's homes will be narrower than their OTT/broadband path through their phone or other 5G device. Although this could create stable conditions for more adaptive rendering via edge computing, this is not yet in widespread use outside the realms of cloud gaming. In the UK, the Technology Innovation Network in partnership with DTG and other industry partners, is investigating use of 5G to distribute media in crowded environments such as stadiums and festivals. Media Objects could be used in this context. https://one4hdd.co.uk/news/

The Ofcom report goes on to add:

> *"Cloud rendering for online gaming is known to be very demanding due to the very low latencies needed between the user's control input (such as firing a gun) and the need to render that scene on the screen. However, cloud rendering for TV and media is less time critical (it's acceptable to wait a few seconds after a button press before the stream switches on the director's commentary). This means that cloud or pre- rendering of simple additions such as extra graphics or an alternative audio version is very straightforward and relatively low-cost for broadcasters.*
>
> *The benefits are that the device for playback can be a very thin client without complex rendering capabilities, or specific new hardware and software capabilities. BBC R&D has published its intention to work with standards like WebAssembly to 'write once and run everywhere' (BBC R&D, 2019), effectively using the cloud to pre-render elements of object-based media that are not software compliant or too processor-intensive for some devices and allowing full device rendering on others.*
>
> *Ultimately however, rendering in the cloud instead of on-device also provides better control of delivery by the broadcaster, improving consumer experience and broadcaster accountability."*

Any discussion of specifications for OBM should therefore allow for the flexible rendering of the final experience, and must allow for renders to;

- be made in production, and for the user to be able to choose the version to watch
- be made in the cloud, where the user could make certain personalisation choices which are re-rendered to a final delivery version
- or allow for all the objects to be delivered to a device where personalised choices are made and objects are played back accordingly.

5. DVB current specifications and previous work related to OBM

5.1. DVB (existing specifications and/or ongoing work)

5.1.1. DVB-I

Overview of DVB-I

The DVB-I solution is a metadata standard for the discovery and delivery of TV services from broadcast or Internet, accompanied by content guides and access to on demand content. Its primary purposes are delivering hybrid TV services to conventional TVs using a familiar user interface and bringing the linear TV experience to other devices such as tablets and mobile phones.

DVB-I stands alongside the existing DVB-T (terrestrial), DVB-S (satellite) and DVB-C (cable) broadcast standards and, for users, the experience of accessing services should be the same, regardless of which delivery channel is used. DVB-I deployments can be stand-alone, using broadband only, or IP delivery can be combined with broadcast to create a single hybrid offering to the user, incorporating services delivered via both methods and making optimal use of the different characteristics of each channel.

The use cases for DVB-I include:

- Deploying a standards based linear TV offering over the Internet, offering equivalent user experience to broadcast TV, but without end users needing broadcast reception.
- Extending broadcast systems with additional IP services, with all services presented in the same user interface, and no user-facing distinction between broadcast and IP systems.
- Supporting the delivery of a linear TV offering over 5G.
- Allowing close integration between broadcast and on demand, for example by supporting restart and selection of on demand content from the linear broadcast UI.
- Offering higher quality versions of services (such as HD, UHD or 8K) to users via IP, where sufficient capacity is not available via the broadcast systems.
- Providing a linear TV user experience on devices other than TVs, including tablets and mobile phones.
- Offering alternative versions of services via IP, with additional accessibility features or regional content.
- Reaching users on TVs without access to DVB-T/C/S reception, for example in IP-only homes.
- Offering fall-back in case of the failure of one delivery channel (for example, a transmitter fire).
- Providing aggregation services for Digital Satellite viewers, allowing targeted channel lists for particular markets and richer content guides.

While taking advantage of the Internet as a distribution platform, DVB-I enables the best features of broadcast TV, that have evolved over decades of television viewing, to be used over IP – such as the integrated channel list, linked applications, the content guide, and simple “lean back” channel selection – while also bringing the advantages of a standardised solution and the scale and cost savings that it offers.

DVB-I Characteristics Overview

To make TV delivery over the Internet equivalent to DVB-T, C and S, the major missing piece from a standards perspective was a service layer, which is the service layer that DVB-I specifies. DVB-I defines the metadata for the discovery of services and services lists over the internet and for retrieving the EPG (Electronic Programme Guide). This metadata is used to signal the services and content that are available, meaning the information used by a TV (or smartphone, tablet, app, etc.) to populate the channel list and the

electronic programme guide. Conceptually, it is equivalent to the SI and PSI information carried in a broadcast MPEG2 Transport Stream.

The metadata describing each DVB-I Service includes one or more Service Instances, these define the delivery parameters for each mechanism by which the Service is available to the client. The delivery can be via DVB-S, DVB-T, DVB-C or over the Internet and the client is free to choose the best option from those available which it is capable of consuming.

Apart from the service discovery aspects, DVB-I built on several established DVB standards which were already available. For instance, IP content delivery relies on the DVB-DASH specification, which was already deployed by many broadcasters, often in conjunction with HbbTV. DVB also added a low latency mode to DVB-DASH, to better support linear live TV services, and this is also supported by DVB-I. In addition, DVB-MABR, the DVB specification on Multicast Adaptive Bitrate (ABR) offers opportunities for broadcasters and network operators to work together to optimise delivery to a large number of receivers simultaneously. This offers a solution enabling DVB-I to scale to very large numbers of users in a cost-efficient manner. There are also mechanisms to extend the DVB-I metadata to signal services delivered via additional networks such as 5G, and it is possible for a service to consist of an application which will present the service itself, or offer other functionality to the user (e.g., a VoD catalogue).

Relationship between DVB-I metadata and broadcast signalling

In a hybrid receiver, DVB-I metadata does not replace conventional broadcast signalling but works alongside it. In a typical installation process a hybrid TV will carry out a channel scan to determine the broadcast services available, the user will then be offered the opportunity to select DVB-I Service Lists. When one or more Service Lists have been selected, the receiver will attempt to match DVB-I Services with (regular) broadcast channels, based on Service name and broadcast parameters, then will present a channel list which combines Services defined by DVB-I with those from regular broadcast. Where a match is found the TV will only show the DVB-I Service including the broadcast channel as one of its Service Instances, if no match is found then the broadcast channel appears in its own right. This means that if a broadcaster does not include its channels in DVB-I Service Lists, or the viewer did not choose a list including those channels, the broadcast channels will still be presented to the user.

The same principles are applied to schedule information: if a (regular) broadcast channel is not matched with a DVB-I Service then its EIT data is used to populate the receiver's EPG, otherwise the DVB-I metadata, which can provide much richer information, will be used instead.

DVB-I service lists

DVB-I Services are presented to users in Service Lists that are XML documents which describe a set of Services curated by a provider who could be a national platform, a single broadcaster or an aggregator. At the top level, Services are defined by their content, with a unique ID, multi-language title and description, logo, language information and target regions. The unique identifier allows the Service to be linked to Service Instances, logical channel numbers and content guides.

Service Lists include additional information to support the delivery of Services. Each Service Instance may also carry a set of Content Attributes which describe the technical parameters of the Service Instance such as codecs and screen size. Service lists also include information about accessibility options.

The information in the Service Instances allows the client to select an instance it has the capability to receive, which is currently available and offers the best quality.

DVB-I service lists support signalling for accessibility features independently from each other via feature specific metadata tagging of media elements or features in linked applications. Beyond this there is no support of more generic object-based signalling of service components on service-list level. In principle, the concept of service-list signalling via XML could be supported (if needed) by extending the XML schema

While it could potentially be possible to create DVB-I service instances for each object-based presentation, allowing users to select which presentation, this functionality is not supported today and would require specification changes. It is also unclear whether the DVB-I service list level is the right place to make such user selections, which might be better made once the user has selected the channel rather than beforehand in the service list.

5.1.2. DVB-TA

DVB-TA defined in bluebook A178 or ETSI 103 752 defines a framework for targeted advertising. Part 1 and Part 3 define the usage of signalling of ad breaks (segmentations of the main content) that may be used to replace or insert an advertisement. Part 2 defines the generic framework for ad insertion including reporting to the ad server of viewer impressions of ads.

The ad slot signalling in part 1 and part 3 is based on the popular Society for Cable and Telecommunication Engineers SCTE 35 specification and is profiled for best usage in MPEG-TS and DVB-DASH. For MPEG-TS DVB-TA part 1 defines additional DVB-DAS descriptors to signal ad slots and constraints are introduced to the fields of markers such as the ‘upid’ type, the command type (splice_insert or time_signal).

DVB-TA part 3 also defines specific recommendations for formatting SCTE 35 and DVB-DASH Events to enable signalling ad breaks. DVB-TA part 1 mainly targets client side ad insertion, where the player/client requests the advertisement to the ad server. DVB-TA part 3 targets both server-side and client side targeted ad insertion. As in the streaming industry server-side ad insertion is currently popular, this is currently the dominant use of DVB-TA part 3. However, DVB-TA part 3 may also be deployed in a similar way as part 1.

DVB-TA part 3 enables signalling ad breaks in a single period DVB-DASH manifest, however output of server side DAI to clients may include multi-period DVB-DASH. While multi-period media presentation descriptions are supported in DVB-DASH, in some cases there are some constraints such as using continuous periods (periods with similar adaptation sets and continuous timing). In practice support for multi-period DVB-DASH may depend on the version and age of the device, recently HbbTV made progress to develop test content with multi-period targeted ad insertion.

The work on DVB-TA has resulted in active collaboration between DVB and SCTE organisations, potentially, DVB could adopt more technologies or specifications from the SCTE organisation.

Currently SCTE 35 is used by DVB for targeted ads only, the SCTE specification does allow for further non-ad metadata (possibly for further objects), but these are not currently used by DVB-TA.

5.1.3. DVB-AVC

The DVB audio and video codecs toolbox (DVB-AVC) is published as DVB BlueBook A001 and ETSI TS 101 154. DVB-AVC has evolved since the first launches of digital TV services to adopt the latest advances in audio and video compression technologies. DVB-AVC targets primarily the usage of multiplexed audio and video bitstreams constituting a contiguous broadcast or streaming service. However, there is nothing preventing the creation of short self-contained audio or video bitstreams, according to the conformance points defined in DVB-AVC, which would resemble the concept of audio and video objects.

The sets of video codec conformance points specified for each video codec generally include luminance resolutions that are smaller than the maximum resolution for the corresponding service class, typically with a quarter-screen image size. For example, for HEVC HDTV video, the lowest defined luminance resolution is 960 x 540. Sub-resolutions are usually upscaled by the receiver when delivered as a broadcast or streaming service’s video bitstream.

Other organisations, e.g. the Open IPTV Forum (OIPF), explicitly extended the sets of conformance points defined in DVB-AVC by specifying video conformance points at so-called sub-picture resolutions, to enable thumbnail video components, for example.

Video format sub-resolutions could in principle be used to carry video objects for rendering on a part of the consumer's screen, but if the intention is to playback multiple video bitstreams concurrently care would be needed in ensuring that receivers have the capability to fulfil such a feature.

The feature of HDR video dynamic mapping (for the display) is realised with any of three optional streams of metadata delivered together with relevant HDR video bitstreams.

The recently completed work item on next-generation video coding (NGVC) adds the VVC and AVS3 video codecs to the DVB toolbox. VVC includes the particular optional features of VVC Subpictures. This feature enables the convenient encoding of video bitstreams that offer personalization and accessibility enhancements by delivering multiple video scenes (objects) within the delivered high-resolution (for example 4K or 8K) video bitstream. The usage of VVC Subpictures for these purposes are described in Annex M of TS 101 154.

The current DVB work item related to video codecs concerns dual-layer video, whereby it is intended that an enhancement layer provides a higher-quality video bitstream when applied to the base layer video bitstream, which represents the backward-compatible baseline quality video bitstream. Potential work on a DVB codec toolbox extension covering dual-layer video based on the corresponding published Commercial Requirements (DVB BlueBook C105) is currently being discussed in TM and SB.

Similarly to video conformance points, the specifications for audio codec usage have evolved from mono and stereo, later adding multi-channel surround sound formats, and more recently adding next-generation audio (NGA). The particular relevance of NGA for OBM is described in clause 6.5.14 above.

DVB-AVC specifies the raw bitstreams for the defined conformance points of codecs that have been adopted by DVB, and how these bitstreams are carried in MPEG-2 TS, and in DVB-DASH (in conjunction with TS 103 285). Services can offer multiple audio bitstreams, along with other components, e.g. subtitles. The receiver requires signalling alongside the audio, video and other bitstreams, in order to render the respective bitstreams correctly for the viewer, and to be able to offer all available options for the user to select. This signalling is referred to as Service Information (SI), and it is defined in the DVB-SI specification (EN 300 468) for services carried in MPEG-2 TS, and in the DVB-DASH specification (TS 103 285) for service carried in the DVB-DASH / CMAF container.

5.1.4. DVB-SUB

Subtitles are important components of media services, since they provide accessibility to audio-visual services for hard-of-hearing viewers, as well as alternative language versions of dialogue. These two different subtitle use cases are signalled to the receiver via dedicated SI signalling [EN 300 468] for services carried in MPEG-2 TS, and in the DASH MPD for content carried in the DVB-DASH / CMAF [TS 103 285] container.

Over the years, DVB has specified three kinds of subtitle formats. The first is the legacy subtitling system known as "EBU teletext", specification EN 300 706, whereby DVB specified their carriage in DVB services in specification EN 300 472.

The next subtitle format to be specified was "bitmap subtitles" (DVB-SUB), in specification EN 300 743. Subtitle texts are carried in the form of bitmap graphics that are rendered on top of the video frames at the receiver. As stated in the scope of EN 300 743, this format can also be used to convey logos or other graphical elements in DVB services, but such usage has not been apparent in DVB services so far. However, DVB-SI [EN 300 468] does facilitate the delivery of graphical icons in either PNG or JPEG format, or via a

URL, in the “image icon descriptor”. The latest edition of DVB-SUB (in 2018) added the progressive-scan format for subtitles, whereby the PNG format was adopted for that. The original DVB-SUB graphics format is based on run-length encoding and accommodates the delivery of two display fields for subtitles, for legacy interlaced rendering.

The most recently specified subtitle format in DVB is a derivative of the W3C SMPTE Timed Text Markup Language, in specification EN 303 560 (DVB-TTML). This subtitle format adopts the web-friendly TTML, i.e. an XML derivative, for text-based subtitles, and the default profile is compatible with both established TTML formats of EBU-TT-D and W3C IMSC1. The latter variant has been adopted by ATSC for subtitling applications.

DVB-SUB also includes the mode of delivering subtitles as text characters, but no usage of this mode is known so far.

DVB-I includes signalling for subtitles on its service discovery processing, i.e., availability and properties of subtitles can be identified by the receiver without the need to tune in the service signal itself, i.e., without parsing the DVB SI or DVB DASH signalling. Please read more on DVB-I at the end of section 6.6.0.

5.1.5. DVB-CSS - DVB CSS (Companion Screens and Services)

In 2015 ETSI published the DVB’s specification for companion screens and Services which consists of 4 parts:

DVB-CSS (Companion Screen & Streams) - Architecture

DVB-CSS (Companion Screen & Streams) - Schema

DVB-CSS (Companion Screens & Streams) - Content ID

DVB-CSS (Companion Screens & Streams) - Discovery

The DVB website for this series of specifications can be found at https://dvb.org/?standard=dvb-companion-screens-and-streams-dvb-css-part-1-concepts-roles-and-overall-architecture

The introduction of the Architecture document explains the purpose and structure of the work;

“Personal, smart devices like tablet computers and smartphones enable new “companion screen” user experiences for broadcast service consumption. These experiences are built on a range of functionality that allow for the connection of the experience on the television and that on the companion device. The DVB CSS suite of specifications can be used to provide these user experiences for the full range of DVB provided broadcast services.”

The Architecture specification is published as ETSI TS 103 286, and can be found here; http://www.etsi.org/deliver/etsi_ts/103200_103299/10328601/01.01.01_60/ts_10328601v010101p.pdf. It describes the key concepts used and the framework within which the other parts of the series of specifications operate and provides an informative descriptive introduction to the interfaces defined in each of the parts.

TS 103 286 architecture specification, provides this additional explanation in its introduction

“The present document, and the framework on which it is based, assumes a network in the home to which devices are connected, that this network is in turn connected to the internet, and that services can be provided from the internet into the home. This allows for the provision of, and use of, services based in the internet that assist the provision of new user experiences. The subsequent parts of this multi-part deliverable provide the details of the interfaces, and provide the further architectural details relevant to the interfaces they define. However, they all work within the generic framework set out within the present document.”

The relevance of these specifications to our study mission on OBM is that it could enable multiple elements and multiple media experiences to be shown on the TV device or on the companion device.

The specification on Content IDs and synchronisation specifies the architecture and protocols for content identification, Timeline Synchronization and Trigger Events for companion screens and streams including:

- the interfaces between the TV Device and the Companion Screen Application:
  - interface for metadata exchange, including content identification;
  - interface for Wall Clock synchronization;
  - interface for Timeline Synchronization;
  - interface for Trigger Events;
- the interface between the Companion Screen Application and the Material Resolution Service (MRS).

  https://dvb.org/?standard=dvb-companion-screens-and-streams-dvb-css-part-2-content-id-and-media-synchronisation

While these specifications were designed to work to allow consumption of elements or media objects to be consumed on multiple physical devices, the concepts are also relevant where multiple media objects are combined and replayed on the same device.

TS 103 286-2 is referenced and used as part of HbbTV (v2.1 and above).

5.2. Previous DVB Commercial Requirements relating to OBM

5.2.1. DVB SSS

In 2018, DVB produced a set of Commercial Requirements for Supplementary Service Streams:

> *"The concept behind the SSS work is one where more than one programme elements are being delivered in different parts of the broadcast or IP delivery chain. Some of this is possible in TV delivery systems today but there are some limitations. These vary based on the type of supplementary streams, where they are carried, and which middleware is used on the receiver. The concept of delivering some program elements or supplementary content separately is becoming more popular, and as many devices now include at least two methods of receiving broadcast or streamed content, the need to establish a reliable and common mechanism to identify, associate, synchronise and re-combine the streams is becoming imperative."*

Although a good number of the commercial requirements were approved by the DVB commercial module, technical work on the requirements was not started. This was mainly due to the lack of member commitment to support the technical work, and a concern that practical use cases were still immature.

5.3. Review of SSS Commercial Requirements work, in light of today's OBM landscape

To better understand if the commercial requirements developed by the SSS group relate to OBM and whether or not they support the OBM use cases discussed in this report, the group studied the requirements and provided comments for each.

1.	Delivery, Identification and Association	Comments by the OBM Study Mission
1.1	The System shall enable the provision for at least 8 supplementary streams to be associated with the Main Programme. Note: this is a system requirement, not an IRD requirement	From the work of the study mission, we have not discovered a method to associate elements from multiple DVB delivery mechanisms delivered via multiple streams. In the case where only a single delivery mechanism is used, more functionality is in some cases possible, but no delivery mechanism has the full capabilities requested in the Commercial Requirements. DVB-I and the DASH manifest helps to solve some of the requirements for OTT, but do not provide a solution for all delivery methods. HbbTV also offers mechanisms to associate streams to a service namely the “Media Synchroniser”.
1.2	The system shall enable delivery of supplementary streams via any DVB delivery method. This shall include both broadcast delivery in the same or different network or multiplex. This shall include internet delivery via DVB DASH Note: this is a system requirement, not an IRD requirement	See above
1.3.	The system shall enable the ability for a supplementary stream or streams to be delivered in advance of the associated broadcast event (seconds, hours, days) Note: this is to allow the IRD to store and replay the SSS with a main programme which may be broadcast at another time (This requirement has been raised by broadcasters in the context of Targeted Advertising, but is also applicable to other supplementary streams)	HbbTV defines a class *Download* to handle supplementary streams that are downloaded in advance. A Download object being made available by the application/oipfDownloadManager embedded object represents a content item that has either been downloaded from a remote server or is in the process of being downloaded. See https://developer.hbbtv.org/references/apis/content-download-apis/the-download-class/ for more details. Additionally, Targeted Advertising has solved many of the issues for advertising, but may not be applicable for other supplementary streams/objects.

1.4.	The system shall provide the ability to carry metadata and signalling to identify and associate supplementary streams. The system shall provide mechanism within the main programme's delivery stream to indicate the location of this metadata. The system shall provide a mechanism for delivery of the metadata via broadcast and/or internet. Note: this also includes a mechanism to identify and associate supplementary streams using only the information contained in the broadcast stream (no internet connection) without requiring the receiver to tune away from the program Note: this may already be possible using existing DVB signalling OR interactive apps	There are methods for most if not all "supplementary streams" which we have also called Media tracks to be delivered in one way or another using DVB (and/or HbbTV) specs. This can be using dedicated DVB specifications such as DVB-SUB, or as video or audio elementary streams. For new media types such as Haptics tracks, there are multiple ways to carry private data. The aspect that is missing is the ability to find all the objects available for a programme, associate them to each other and create logical and functional experiences that allows you to use combinations of objects. In other words it's SI and metadata not media transport mechanisms that are missing
1.5.	The association of program elements shall allow relationships and dependencies between supplementary streams to be indicated Examples of relationships would include Replacement - where a supplementary stream replaces part or all of the Main Program e.g. a replacement audio for a program Enhancement – where a supplementary stream provides an enhancement or additional element in addition to the main programme e.g. director's commentary, or signing for the deaf Examples of Dependencies would include Valid combinations of main and supplementary streams e.g. English AD track is valid only when the main audio is English	Preselections in NGA provide information on how audio streams can be re-combined, including which combinations are valid. For non-audio presentations, this mechanism is missing HbbTV apps and/or Media Synchroniser offers some functionality

1.6.	The association of program elements shall support presentations which combine one or more supplementary streams with the Main Programme. The system should include a method to indicate which combinations are valid. A valid presentation might be Main Programme Video with French SSS Audio and French SSS Audio description Note: re-use of existing mechanisms is encouraged.	This CR has been solved in individual apps but there is no standardised way to signal valid combinations of objects except the limited functionality in the DASH manifest. To make OBM viable in the market it is imperative that valid and invalid combinations of objects can be signalled and selected by devices.
1.7.	The system shall allow a Simple description of the information carried in the supplementary stream to be identified prior to selection that stream e.g. is it audio/video/picture etc. (Note: this could be considered to be comparable to the information conveyed in the EIT component descriptor defined in EN 300 468) The system should also allow a mechanism to carry more detailed information about the supplementary stream to be identified prior to selection that stream e.g. resolution, Note: some signalling is already possible, but analysis of the use-cases needed to verify if this signalling can be used to fully or partially fulfil these requirements	DVB-I service selection does have a mechanism to identify the versions of the service available and to guide the decoder only to the content that needs to be decoded but it does not have full object support.
2.	Programmatic elements	Comments by the OBM Study Mission
2.1.	The system shall support the delivery, identification and association of all types of Supplementary Service Streams These include but are not limited to any combination of streams such as; 2.1.1. Video (streams, objects, elements, enhancement layers) 2.1.2. Audio (streams, objects, elements, enhancement layers) 2.1.3. Still images e.g. photos or logos, graphical elements 2.1.4. Subtitles (streams) 2.1.5. Data and Metadata e.g. information on the actors or locations, metadata on the items or characters in the scene etc. 2.1.6. Applications play along broadcast games or voting application. (Note: 2.1.6 may be solved by AIT, but TM should verify if this method is suitable)	Applications may be complex as the application itself may be used to select and play back objects. In addition to the list of objects in the SSS Commercial Requirements, there have been advances in the types of elements that may now be in use, Haptic output (a track that is played in time with the content), 2D and 3D elements, Signing. Branching Narrative metadata. Proprietary applications now exist to deliver many if not all of these objects in the context of a program. Some operating systems already contain capabilities beyond what is possible using standards based solutions (for example the Sky "EntOS" operating system)

3.	Synchronisation	Comments by the OBM Study Mission
3.1.	The System shall enable a mechanism for the synchronisation of all valid combinations of programme element or supplementary stream delivered by broadcast and/or IP	The CSS specification from DVB has now been (partially) adopted as optional by HbbTV 2.03 clause 13 and is in use for media sync. Primarily this was designed for companion screens, but is fully capable of being used on the same device as well. It is anticipated that the current CSS specification could be utilised for OBM and that no further specifications would be needed for synchronisation.
3.2.	The specification shall enable associated Supplementary Service Streams to be synchronised with the Main Programme presented or selected on the TV Device.	See the comments for 3.1
3.3.	Broadcast Synchronization Accuracy Levels The specification shall define mechanisms for the following synchronization accuracies so as to cope with different use cases and needs: · Service/Event accuracy · ~2 seconds accuracy · Lip sync accuracy · Frame accurate This requirement shall be met when streams are received through broadcast or broadband delivery The broadcaster can provide different levels of accuracy for different supplementary streams for a service. Note: TM should advise which level of accuracy would be practical or technically feasible Note: The Synchronisation accuracies above match those of DVB-CSS Note: The accuracy levels are in-line with COS, but need to be checked with the published CSS spec	The CSS specification already covers the 3-4 main levels of accuracy needed, but as mentioned in the Commercial requirement, it is not a requirement in today's devices to implement. In addition to the levels of accuracy, it is conceivable that some objects may need a level of viability or optionality in the playback sync (e.g., signing, where the user may prefer a longer or shorter delay from action to signing.
3.4.	Blank	

3.5.	It shall be possible for the synchronization mechanism to be robust against re-multiplexing, transcoding, etc. taking place during TV delivery (both within and across TV Platforms). Note: CSS mechanisms are believed to already fulfil this requirement.	CSS has been shown to be robust against delivery mechanisms and varying delays.
3.6.	The specification shall enable an IRD to receive the synchronization information within a maximum of 2 seconds of the service selection. Note: Tuning and decoding has some end-user noticeable delay during the service selection, during which it is not expected to be able to synchronise against the main programme	Service selection is probably the wrong metric here, as selection to reception time can vary greatly due to disparate servers, DRM, and other factors, this CR should therefore be interpreted as the sync information being available within 2 seconds of the content being received and decoded. The SM group is not aware of services that send sync information in DVB services today.
3.7.	The specification shall allow the IRD to synchronize against Main Programme any time during an Event (e.g. a TV programme)	This is believed to be covered by the CSS specification.
3.8.	CM-SSS requires the ability to synchronise anywhere along the Main Programme's Timeline. The system should allow the synchronisation to be maintained across TV programme or event boundaries on the same service	The group is not aware of services that send sync information in DVB services today. The CR is still valid for OBM systems which would need to maintain sync
3.9.	The specification should allow for the IRD, once synchronised, to switch or add a supplementary stream to the main program with minimum disturbance to the user experience Note: the SSS CRs are only concerned with the system that allows for synchronisation, identification and association, and do not address the complexity of minimising disturbance when switching audio, video and graphics on the IRD once the streams are available.	For NGA today, timestamps are used for content that has passed through the same encoder, and these timestamps can be used to sync audio that arrives through different paths. This works for related content from the same source, but not for content encoded elsewhere. CSS does have this capability, but is not widely used nor implemented
4.	Combination	Comments by the OBM Study Mission
4.1.	Note: Specific requirements on combining the contents of Supplementary Streams together to form specific presentations of picture, video, audio and graphics are not included in these requirements but may be part of a subsequent phase of this work	This SM has discussed the need for OBM systems to know which combinations of objects can be used to make a valid playback selection.
5.	Improved Accessibility and Usability	Comments by the OBM Study Mission

5.1.	The System shall enable users to easily select combinations of main and supplementary streams to create a presentation. (User interface and user control is out of scope of these requirements.	The importance of usability and accessibility in the selection of objects is a primary concern, systems should allow automatic selections and simplified selections of a "preselection" as in NGA where possible
5.2.	The system shall include methods that can be used to prevent adverse effect on accessibility of the final presentation of the programme including supplementary streams (e.g. multiple language commentaries played simultaneously). e.g. include accessibility criteria in the dependencies and valid combinations discussed in CR 1.5	The SM agrees that OBM presentations must have the ability to prevent the presentation of a program being compromised by invalid object selections. This implies some level of communicating valid combinations. In Audio this is already handled by the "preselection" mechanism
6.	Personalisation	Comments by the OBM Study Mission
6.1.	Provisional: The System shall enable users to select combinations of main and supplementary streams to create a personalised presentation, created from a combination of the available supplementary streams.	Whilst the concept of "pre-set" presentations providing a specific combination of content is established in NGA, there is no overarching standard to allow simple pre-sets to deliver a combination of content across audio, video and data streams.
6.2.	The system shall allow personalised selections of Main Programme and supplementary streams based on user input, and/or user preferences known to the decoder (e.g. language, or Accessibility preferences)	If a user has selected a preferred language, or a preference for subtitles to be on, this should be delivered (where available) for any source or content. HbbTV is working to enable this through their Accessibility Framework.
7.	Hybrid delivery	Comments by the OBM Study Mission
7.1.	Provisional: The System shall enable the flexible hybrid delivery of Supplementary streams, the synchronisation requirements also apply in all hybrid delivery cases Note: network latencies will be different for the different delivery mechanisms, Synchronisation information should be provided but it is understood that synchronisation may not always be possible due to buffering and relative delays in each path. The system should however provide all the synchronisation information to allow capable devices (with delay /buffering capabilities) to sync the streams	Not discussed by the group

7.2.	The primary and secondary mechanisms may both be broadcast TS, or DVB-DASH or any combination of the two, and may be streamed, and/or pre-loaded on the device.	Not discussed by the group
8.	Encryption	Comments by the OBM Study Mission
8.1.	Provisional: The system should not prevent encrypted or unencrypted supplementary service streams to be supported. The system should not prevent these (encrypted or unencrypted) supplementary streams to be consumed together with the Main Programme where the main programme itself is encrypted or unencrypted. The system should not prevent the supplementary stream from utilising a different Conditional Access system or different access rights from the main programme Note: the system should take into account any practical constraints in the encryption-decryption (sub-)system(s) to ensure practical deploy-ability of a secured Rich Media	Not discussed by the group
8.2.	Provisional: The system shall allow the SSS streams to be hidden from scan	
9.	Companion screens	Comments by the OBM Study Mission
9.1.	Provisional: The System should be compatible with the existing DVB CSS specification.	Not discussed by the group
9.2.	Provisional: Timelines should be reused wherever possible, and stream identification methods should be re-used or should complement each other. If changes are necessary to address the hybrid delivery, they should be kept as small as possible.	Not discussed by the group

6. Work of other groups and Standards Development Organisations (SDOs) on OBM

It has become clear to the group that many other SDOs are either already working in the area, and have specifications which support some aspects of OBM, or have groups who have or are planning to work on similar topics. As the topic is large, it touches upon standards for web protocols, asset management specifications, and immersive audio and video specifications, streaming protocols as well as mark-up languages of many kinds.

In the following section we will discuss the elements of current standards which we consider to be relevant for OBM.

6.1. SMPTE and IMF-UG

SMPTE is driven by the needs of the industry and, with respect to “Object Based Media”, is engaged with the needs for the Movie and Broadcast Industries in the use of Objects under the “Interoperable Mastering Format” – IMF. Please refer to https://www.smpte.org/standards/st2067.

IMF is a file-based media format that simplifies the delivery and storage of audio-visual masters intended for multiple territories and platforms. It is particularly well-adapted for delivery to today’s global content platforms. IMF works with any finished audio-visual masters, including long-form movies, episodic content, advertisements, short-form content, etc. At its simplest, it contains a video track, an audio track and an XML document (Composition Playlist [CPL]) which describes how to play the video and audio together. The CPL is the key to how IMF works, and can be used to play back any amount of media and metadata in any order.

IMF is extensible, uses proven technologies and benefits from many open source and commercial implementations. Work is taking place in the IMF User Group and SMPTE to extend the IMF format to carry multiple video tracks and time-variant metadata which will further extend its usefulness.

IMF allows the reuse of media in multiple versions of a piece of content which simplifies the storage and distribution of that content, however it is designed to be used for handling complete programmes. To carry multiple narrative pathways through a story would require a CPL to be created for each of those pathways.

6.1.1. Work of the IMF User Group

The IMF UG brings together content owners, service providers, retailers and equipment/software vendors to enhance and promote the use of IMF globally, across domains of applications. The IMF UG discusses technical operational issues that arise in practical implementation, conducts and cooperates with other industry bodies and standards organisations on interoperability testing, develops best practices, and seeks to broaden the awareness of IMF.

6.2. ITU-R

ITU-R Study Group 6 is working on recommendations and reports on the development of object based componentized media.

Report ITU-R BT.2400 Annex 2 describes a potential content exchange format that is capable of supplying multiple versions of non-live programmes to a global platform for broadcasting. It describes both generic componentized content workflows and the Society of Motion Picture and Television Engineers (SMPTE) Interoperable Media Format (IMF). Because of territorial, regulatory and rights requirements, multiple versions of content is often required and different platforms may require technical variants to accommodate a

variety of display formats and capabilities. The ITU sees the use of componentization as a way to address the international exchange of content and a way to minimise the amount of data storage and therefore the environmental impact of media creation and exchange. The Report is strengthened by Recommendation ITU-R BT.2153 *The use of componentized workflows for the exchange of non-live television programmes*.

Report ITU-R BT.2524 provides information on object based media production and how a componentized system can be used to package and exchange object based media programme material. This report also suggests how accessibility services can be described as objects in any media package and suggests objects can be described as components.

The Report goes on to suggest that to realise the potential of object-based media, it is important that a set of standardised rules are developed. Rules will be required to define how the data needed to describe how an object is packaged and adapted by the applications used to process and present the content to the end user. This area is under study to produce recommendations for the production and international exchange of object related data and terminology.

6.3. HbbTV

HbbTV references the DVB toolbox and includes the capabilities for media encoding and delivery that are provided by DVB. The HbbTV specification includes conditional mandatory support for Next Generation Audio use-cases that enable viewer interaction for personalising the audio experience. Next Generation Audio streams received via broadband that allow personalised rendering can be controlled by an interface between HbbTV application and receiver to choose one of the available Audio Preselections (rendering pre-sets) delivered within the NGA stream.

Feedback from the HbbTV association stated that, beside the NGA use-cases for object-based audio there is not much support for other object-based media usage in HbbTV, today. HbbTV provides an opportunity to propose new commercial requirements to their members, starting at the beginning of every new specification cycle. Support for new object-based media use-cases beyond the currently supported ones for NGA can be proposed.

6.4. MPEG Systems Technologies for Object Based Media

6.4.1. Object Based Representations in MPEG for 3D environments

Object Based Media was already introduced in MPEG-4 which provides an object based model in MPEG Binary Format for Scenes (BIFS). However, the standard did not reach sufficient traction and market adoption. Current emerging work in MPEG is on scene description MPEG-I, ISO/IEC 23090-14 that basically enables descriptions of 3D environments with different BIFS nodes enabling insertion of different media at different places in the scene graph for 3D rendering. MPEG Scene description and BIFS are mainly relevant for 3D environment rendering and work is done in collaboration with Khronos. In the 3D environment light sources, video texture sources, 3D Meshes can be inserted based on MPEG defined representations.

6.4.2. CMAF as a model for object based media

The Common Media Application Format defined in ISO/IEC 23000-19 is rapidly becoming one of the popular formats for streaming due to adoption in popular profiles from DVB, DASH-IF and support in HTTP live streaming. Recent addition in the 3rd Edition includes support for timed metadata tracks that could enable various use cases for object based media. In CMAF different switching sets can be labelled for their usage using the kind box and a track role such as those specified in MPEG-DASH or in HTML 5. This

enables labelling of the groupings for specific usage such as viewpoints or different accessibility features such as additional video descriptions. DVB may be able to contribute to consistent implementation and support of different track roles defined in MPEG-DASH.

CMAF Switching sets or aligned switching sets may be used to implement different objects. The relationship between objects can be defined using a media presentation description or using specific identifiers in the media tracks such as in the kind box.

6.4.3. MPEG-DASH for object based media

MPEG-DASH published the 5th edition in 2022, the main updates in this edition include the CMAF binding (CMAF profile for DASH). Also an event and timed metadata processing model that is already (partially) supported in DVB-DASH that can enable object based media use cases is included. For example timed metadata events can be used to carry interactive information relating to the media presentation. The client processing model defines how and when timed metadata should be processed by a player and passed to an application.

The more recent advances in amendment 1 to MPEG-DASH 5th edition include support for some of the use cases from this study. One is a mechanism for non-linear playback explicitly adding signalling for stitching different periods or pieces of content based on user input or other factors, also using an EventStream. In addition support for insertion of (pre-roll) or (mid-roll) in presentations is supported via new signalling (using EventStream).

MPEG has developed extensions for branching narratives with alternative MPD event in amendment 1 of the 5th edition (to be integrated in the 6th edition). It is an event for MPEG-DASH with scheme uri urn:mpeg:dash:event:alternative:2022 it enables switching to a different mpd at a certain point in the content. In addition, after playing the alternative MPD the player switches back to the original content. The DASH specification also gives way to use this to insert mid rolls.

In addition, in annex K an event with scheme uri "urn:mpeg:dash:nonlinearplayback:2020" is defined to enable a nonlinear storyline, combining different content. This can all be in the media presentation description but this can only work for on demand content so it covers only a subset of DVB use cases. But the nonlinear playback is another alternative for implementing branching narrative.

While these approaches have been adopted in MPEG-DASH amendment to 5th edition, it is unclear if they are adopted in practice or in application standards and a review of the maturity of this work would be needed.

6.4.4. MPEG File Format for object based media

MPEG systems group (ISO/IEC JTC1 SC29 WG3) maintains and develops different specifications using the ISO Base Media File Format defined in ISO/IEC 14496-12. MPEG-B ISO/IEC 23001 series includes file format specifications for metadata and other file format data. In particular interest could be for object based media:

A) MPEG-B part 10 timed metadata metrics in ISO BMFF. This spec may be useful to add metrics in the stream, it mainly focuses on quality metrics such as PSNR, SSIM etc. It uses carriage in ISO BMFF, potentially making it easy to integrate in DVB-DASH.
B) MPEG-B part 15 carriage of web resources in ISO BMFF. This spec defines embedding of HTML and JavaScript resources in ISO BMFF, and describes a model for synchronised rendering. This could potentially be used to insert small snippets of web content in a broadcast stream that is sent using MP4 such as used in DVB-DASH or DVB-MABR
C) MPEG-B part 18 event message track in ISO BMFF. This is a generalised format similar to WebVTT in MP4 (MPEG-4 part 30) that allows carriage of event messages in separate mpeg tracks.

This format can be useful for example when 3rd party metadata from a different source is integrated. It can also be integrated in DVB technologies similarly as timed text or subtitle sets using a separate adaptation set.

DVB could do additional work on signalling this type of content in the Media presentation. This may need some coordination with other organisations like MPEG or DASH-IF.

6.5. W3C

W3C develops standards for web browser APIs, data formats and schemas. Some web standards, notably the HTML specification, are developed outside W3C, in WHATWG, although there is close collaboration between the two bodies.

W3C has a number of media related groups, but does not currently have a dedicated workstream on object-based media. The W3C Media & Entertainment Interest Group would be the natural home for technical discussion of object-based media use cases and support in web browsers and web specifications, and work on gap analysis. This group would welcome further discussion and collaboration with DVB members.

Several of the technologies W3C has developed, and is currently working on, are relevant, including Media Source Extensions, Web Codecs, Web GPU, and Web Audio. The following sections describe some of these in more detail.

6.5.1. HTML video and audio elements

HTML enables web pages to be composed of text, images, video and audio, and Cascading Style Sheets (CSS) enables responsive positioning and layout of elements in a page (e.g., to adapt to different screen resolutions), as well as animations and transition effects.

The HTML <video> and <audio> elements allow web developers to include video and audio content in web pages. The browser provides default player controls (play, pause, seek, audio volume, etc.).

HTML elements, including text and images can be overlaid over an HTML <video> element. This capability is often used to show subtitles or captions and custom player controls. For object-based media applications, it can also be used to show interactive elements related to the video as layered content. However, this ability to overlay content may be more limited on TV devices than on desktop and mobile.

Layered content can be synchronised with media playback, for example by responding to text track cue events, or by querying the current playback time position, with some limitations. A report from the W3C Media & Entertainment Interest group, Requirements for Media Timed Events, section 5.3 discusses these in more detail.

6.5.2. Media Source Extensions

Media Source Extensions (MSE) allows web applications to construct media streams for playback. It enables HTTP adaptive media streaming (e.g., DASH) in browsers. MSE was published as a W3C Recommendation in 2016.

The Media Working Group is currently developing the next iteration of MSE, known as MSE v2, which is planned to include several new features:

Codec switching – enables support for codec or bytestream transitions, e.g., to enable client-side ad insertion, or branching narratives in an object-based media context, if branches may use different media encodings.

Worker support – improves web page performance and media streaming performance (including start-up time and reducing stalling) by allowing MSE operations to be performed in a Web Worker, which reduces load on the main Window context, which is responsible for handling web page user interface updates.

Improved memory management – The proposed ManagedMediaSource API allows the browser to have more control over media buffering and when media segments are appended. The goal is to improve streaming media playback support on constrained capability devices, as well as reduce power usage, by allowing the browser to react to changes in available memory and networking capabilities.

Web applications can use the HTML Video element together with Media Source Extensions to support OBM branching narratives. In principle, this can be done with appropriately placed splice points in the media, and using the MSE API to manage the buffered media. However, this approach may not support instantaneous switching at arbitrary time points.

The workaround for this requires pre-buffering of multiple video elements and using JavaScript to replace the video element in the web page at the right time. This itself has some limitations, as JavaScript execution is asynchronous to media playback, so frame accurate transitions may be difficult to achieve.

Further investigation into the strengths and limitations of these approaches is recommended. In addition, the new features being introduced in Media Source Extensions v2 should be evaluated in an OBM context, such as the proposed ManagedMediaSource API.

6.5.3. Media Capabilities

The Media Capabilities API allows web applications to query whether the browser and device can render or record a given audio or video media format, for streaming and WebRTC. It provides precise results compared to previous APIs, such as isTypeSupported in Media Source Extensions, and the HTML Video element canPlayType method. Queries can include codec string, resolution, bitrate, frame rate, and HDR parameters, and the result indicates whether decoding or encoding is supported, whether playback is expected to be smooth, and if the browser implementation considers it to be supported in a way that is power efficient.

The decode API is supported in all major browsers. One proposed addition to this API is the ability to query the ability to seamlessly transition between configurations, e.g., for client-side content insertion.

6.5.4. WebCodecs

The WebCodecs API provides low-level access to the browser's audio and video decoders and encoders (which may be software or hardware). The use cases include very low latency streaming (such as for cloud gaming), media editing, and transcoding. The WebCodecs Codec Registry defines codec-specific settings.

For media playback and rendering, the inputs to WebCodecs are encoded media bitstreams, and the outputs are VideoFrame and AudioFrame objects containing the decoded media. Video frames can be rendered using the HTML Canvas API, or processed using WebGPU or WebGL.

Audio can be rendered using an AudioWorklet (see section 5.7.5 for more details).

WebCodecs can potentially be used to support object-based media use cases:

Branching narratives – Using WebCodecs, the web application fully controls the buffering and rendering of media through JavaScript. This could enable branching at arbitrary time points in response to user input.

However, it moves the complexity of media playback from the browser to the JavaScript layer (for container parsing, media buffering, seeking, playback timing, etc.)

Layered composition – Web applications can render WebCodecs video frames using the HTML Canvas API. This would also allow multiple video layers to be combined, text and graphics overlays to be added, or visual effects (fades, transitions, etc.) to be applied in real time.

6.5.5. Web Audio API

The Audio Working Group develops and maintains the Web Audio API, which allows web applications to process and synthesise audio, to support use cases beyond basic streaming audio playback, such as for games or interactive applications.

The specification reached W3C Recommendation status in 2021. Recent additions include AudioWorklet, which allows web developers to supply scripts (such as JavaScript or WebAssembly code) to process audio on a dedicated audio rendering thread, with low latency. This ensures synchronous execution of the script code with other Web Audio processing.

6.5.6. WebGPU

The WebGPU Working Group develops the WebGPU API, which provides low-level GPU access to web applications. The API is based on modern OS-level graphics APIs such as DirectX 12, Vulkan, and Metal. The group is also producing the WebGPU Shading Language specification, to enable web developers to create shader programs that run on the GPU.

In an OBM context, this can be used to process video frames, e.g., from the WebCodecs API, apply visual effects or compose graphics.

6.5.7. Web Transport

The Web Transport Working Group is developing the Web Transport specification. This provides a browser API to support the IETF Web Transport network protocol. Web Transport enables low latency media delivery using unreliable transport, to avoid head-of-line blocking issues inherent with TCP delivery, in a way that is compatible with the web security model. This may be useful for object based media experiences and low latency streaming to browser (and other) clients.

6.5.8. High dynamic range and wide colour gamut support

Work is ongoing in the W3C Colour on the Web Community Group to develop proposals to enable HDR support (including HLG and PQ formats) in the HTML Canvas API, which is currently limited to 8-bit graphics. This would allow web applications to render and compose in HDR, such as video frame overlays, and to match colours between video and web content such as text and graphics. This includes content overlaid over or composited with the HDR video, as well as content in the surrounding web page.

Future work may include extending WebCodecs and WebGPU to support HDR.

6.5.9. PNG graphics

The PNG Working Group is producing an update of the PNG image format specification, including support for HDR images and HDR metadata.

6.5.10. Timed text and timed metadata

The Timed Text Working group develops and maintains the Timed Text Markup Language, (TTML), TTML Profiles for Internet Media Subtitles and Captions (IMSC), and Web Video Text Tracks Format and API (WebVTT).

TTML and IMSC are XML based formats that describe timed text that is associated with timing information. In addition to delivery of subtitles and captions, the TTML timing model could be interesting to consider as a way to describe synchronised content in object-based media.

WebVTT supports arbitrary timed metadata, e.g., through the use of JSON objects in WebVTT documents. This can be used with the HTML TextTrack API to trigger media-timed events to drive synchronised object-based media experiences. This approach has timing limitations, as discussed in section 5.7.1.

The Media & Entertainment Interest Group developed a DataCue API proposal to give more direct support for timed metadata, rather than relying on the WebVTT cue API which is limited to providing data in string form. This is also discussed in the Requirements for Media Timed Events document.

Note that both TTML and/or WebVTT may be encapsulated in MP4 or fragmented MP4 (e.g. CMAF) based on ISO/IEC 14496-30, making these two formats easy to integrate into streaming or other MP4 based workflows.

6.5.11. Synchronized Multimedia Integration Language (SMIL)

SMIL is an XML-based language for interactive multimedia presentations. It includes timing, layout, animations, visual transitions, and media embedding. The SMIL 3.0 W3C Recommendation was published in 2008.

SMIL could be interesting to evaluate from an object-based media perspective, as a way to describe and convey OBM media experiences.

6.5.12. Metadata

W3C has standardised a number of generic data formats, including XML and JSON-LD.

Schema.org is a general purpose vocabulary for structured data, for use by search engines, developed in a W3C Community Group. The vocabulary includes a number of terms that describe media content, such as CreativeWork, Movie, MusicRecording, TVSeries, AudioObject, ImageObject, and VideoObject.

These standards could be considered as a way to describe the media content, and to describe and reference transmedia objects associated with the main content.

6.5.13. Future work

The W3C Media & Entertainment Interest Group would be the natural home for technical discussion of object-based media use cases and support in web browsers and web specifications, and work on gap analysis. This group would welcome further discussion and collaboration with DVB members.

6.5.14. Next generation audio

Dolby, Fraunhofer, and BBC have collaborated on a set of use cases and a proposal for browser support for next generation audio codecs, including AC-4 and MPEG-H. This proposal was presented at W3C's annual technical plenary meetings in September 2022. If browser vendors are supportive, the next phase of work would be to develop prototype browser integration and develop an API specification.

6.6. AI4ME

The following text is derived from an article in DVB Scene edition 62

https://dvb.org/news/dvb-scene-62-looking-back-and-looking-forward-on-dvbs-30th-anniversary/

Delivering the future of interactive and personalised media at scale

Graham Thomas, Rajiv Ramdhany, BBC R&D, UK

Adrian Hilton, University of Surrey, UK

Nick Race, University of Lancaster, UK

The media sector is in the midst of a transition from broadcast infrastructure and services to IP. In addition to the challenge of maintaining the scale and resilience of broadcast, we also need to meet audience expectations that internet-delivered content can be personalised and customisable. This includes assembling content to meet particular user needs (such as accessibility), device characteristics (like screen size), audience interests and context (such as the order, duration or prominence of certain sections or overlays), or business needs (such as advert placement or accounting for rights agreements for specific regions).

Substantial research has already been undertaken on object-based media which lays the foundation for this, but further capabilities are needed to make it possible to create and distribute truly flexible content on a large scale.

The AI4ME Prosperity Partnership is laying the foundations for building those capabilities, by addressing key challenges in production and delivery. The partnership is led from the academic side by the University of Surrey and from the industrial side by the BBC, with Lancaster University bringing specialist expertise in software-defined networked distributed compute.

The University of Surrey is developing AI-powered techniques to allow audio and video to be separated into objects (such as individual audio tracks and distinct 'layers' from a video scene) and create metadata describing the scene.

There is a significant overhead to creating object-based content in current production processes, as it is geared towards creating 'finished' programmes rather than the individual components needed to produce flexible content. BBC is investigating potential use cases and production tooling that would be needed to support them, building on the content analysis work from the Surrey team.

Lancaster University is addressing the challenge of delivering and rendering customised content, by working with the BBC to develop approaches to intelligently distribute the processing and data through the delivery network, making optimal use of resources in the cloud, edge-compute nodes, and the audience's networks and devices. To render customised content at present, high-end audience devices or large amounts of cloud computing resource are required, making it expensive and impractical to reach large audiences.

To deliver highly interactive and personalised media, the infrastructure should be capable of pushing composition and rendering processes further down the delivery network closer to the user and re-compose/re-render content when required. But whilst future media delivery systems must provide this level of

*placement flexibility, they will still be bound by the tenets of **access-universality, scalability** and **sustainability**.*

To enable rich new features universally for audiences, AI4ME is looking into approaches to build flexible media applications that can adapt to the end-device capabilities and selectively offload compute-intensive operations to the cloud or edge. For cost-effective and environmentally sustainable delivery, strategies that allow sharing of compute resources e.g. GPUs and generated results, will enable a deployment of media-delivery services to scale sub-linearly with increasing audience size. Since media delivery systems will likely span multiple heterogeneous compute targets and networks, intelligent provisioning and orchestration across these are required to fulfil delivery business objectives such as Quality of Experience (QoE) thresholds, and cost/energy KPIs. As well as addressing the above challenges, the BBC and Lancaster University are exploring novel real-time adaptation and optimisation techniques to enable delivery systems to gracefully degrade to maintain user QoE and cost/energy budgets as operating conditions and audience sizes change. With these new capabilities, the delivery infrastructure and the content can both flex to meet the audience needs and content-provider's distribution objectives.

6.7. AVS SMT

The Smart Media Transport protocol created and published by AVS (Audio/Video Working Group of China) is designed for future-proofing various applications, and provides personalised services based on a user's profile. One service could offer multiple presentations which then use different objects.

https://standard.avs.org.cn/avs3_download/en_index.asp Information Technology - Intelligent Media Coding (AVS3) Part 6: Intelligent Media Format (T/AI 109.6-2022)

Encapsulation format: SMT supports flexible object organisation to support dynamic service configuration. SMT provides a media-aware data model for content encapsulation, and the relationship among different objects are indicated.

Heterogeneous network: Objects of one service could be delivered over different kinds of networks, e.g., some objects over a broadcast network and others over broadband. And objects from different networks could be received by the same terminal and presented together. In this case, common and popular content could be delivered over a broadcast or multicast network, and personalised contents could be retrieved based on user requirements.

Various types of QoS: Objects within one service can have different QoS levels.

Context-aware data protection mechanism for reliable transmission: Objects are protected adaptively based on their priority.

Network clock synchronisation: Objects delivered among different networks could be synchronised.

Interaction between multiple devices: Objects of one service could be presented across multiple devices, and users could select and switch them smoothly.

Media synchronisation control: Different objects could be presented synchronised on the same device, or different devices.

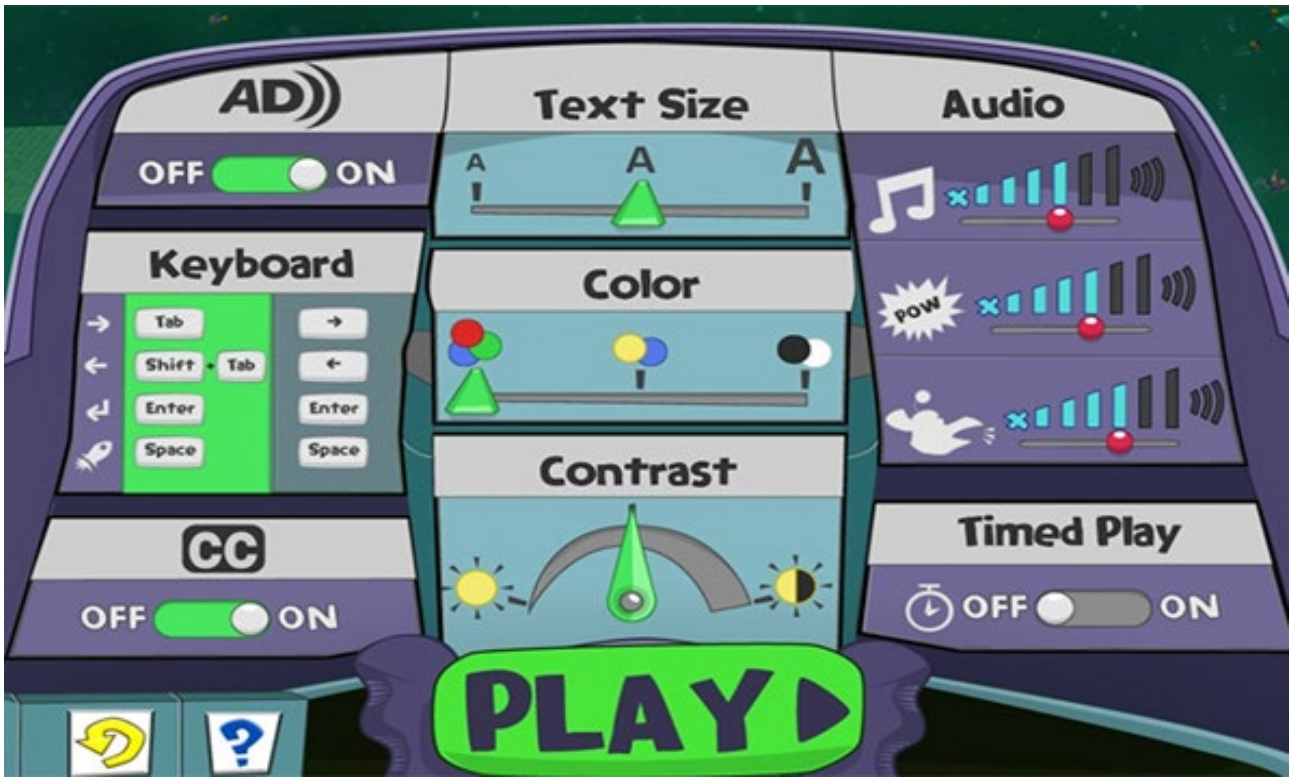
One service could be constructed by various objects, they could be added, replaced, removed and modified flexibly, delivered over same or different networks dynamically, and presented over same or different devices synchronised. So the users could consume the service in various ways based on their preference. This is also the motivation of OBM.

7. Commercial use cases for OBM

During the course of the study mission the group received presentations outlining many use-cases that have been either trialled, or are in use on certain platforms today.

UC1: Advanced and mixed accessibility use-cases

Stornaway were asked by Channel 5 to produce an object based media experience with different experiences based on user needs, e.g., a version for people with learning difficulties, ADHD, or dyslexia, or multiple needs in combination. The experience enables a combination of accessibility features to be enabled simultaneously, and could include large size subtitles, dialogue enhancement, high contrast, and reduced background sounds or music based on narrative.



Picture provided by Stornaway.io, and may not be representative of the UI seen by users/viewers

Many of the individual accessibility elements are well supported today and can already be delivered. The novelty presented by this use-case is the combinations/experience/presentations that can be selected.

For the objects that are delivered by the content provider an app could be used to communicate which of the various options could be enabled simultaneously or which combinations of which objects are recommended for certain user/accessibility groups.

This would rely on such metadata being exchanged between the content provider and the user via the app, in a proprietary way.

UC2: Branching narratives in linear programming (Peacock /Netflix)

Extracts from an article by TechCrunch

https://techcrunch.com/2022/10/14/peacock-experiments-with-interactive-scenes-to-give-fans-a-real-housewives-deep-dive/?guccounter=1

Image Credits: Peacock

"NBCUniversal's Peacock has trialed a new interactive video feature that will let Peacock Premium subscribers delve into extended clips, including extra footage and interviews, from within an episode of "The Real Housewives."

Season 3 of the hit series is set to premiere in 2023 and will be the first to have the new feature. At launch, premium subscribers can interact with three episodes of "The Real Housewives Ultimate Girls Trip," with the option to watch extended clips and exclusive interviews.

As viewers watch the show, they will be prompted to get their remote ready for an "interactive scene" and have around 15 seconds to decide if they want to watch extended footage of that scene as well as exclusive interviews, which go deeper into what characters were thinking during that particular scene.

"Combining culture-defining content with cutting-edge innovation, this experience is about giving fans the choice to dive deeper into the most dramatic 'Housewives' moments on their own terms," John Jelley, Senior Vice President, Product & UX, Peacock and Direct-to-Consumer, NBCUniversal, said in a statement.

Viewers are not required to select anything and can continue to the main story, though anyone who wants to can watch a director's cut-like version of the episode. "This is an opt-in experience. It's not something where we're forcing everyone to change the way they're watching their favorite show," Jelley told TechCrunch. They also have the option to go back to the main story at any time.

This new "unique experience" will hopefully get new fans to the platform, Jelley added. "Now that we're an exclusive streaming home for Bravo content, it's an opportunity to show Bravo fans that Peacock is a great place to come for this type of content," he said.

Because "The Real Housewives Ultimate Girls Trip" is a Peacock original, the team at Bravo worked in close partnership with Peacock to film scenes specifically for the interactive feature exclusive to the platform. For instance, Peacock is exploring giving viewers the option to watch what's happening in the next room during a scene, Jelley told us.

In a demo with TechCrunch, Jelley gave us a sneak peek of the feature. We watched the coffee reading scene in the second episode of season 2 when Dorinda invited the other cast members to her famous Bluestone Manor. Interactive scenes included extra footage of Dorinda blowing up at Brandi and exclusive interviews from Eva, Tamra and Brandi.

Peacock continues to invest in interactive features to engage viewers further. In May, the company announced "Catch Up with Key Plays," which lets English Premier League viewers stream highlights without disrupting the game. Peacock recently launched a "Halloween Nightmare Game," an interactive virtual escape room game that makes the viewer click on different objects on the screen, e.g., reality TV where you can focus on one contestant."

Extracts from an article by Advanced Television

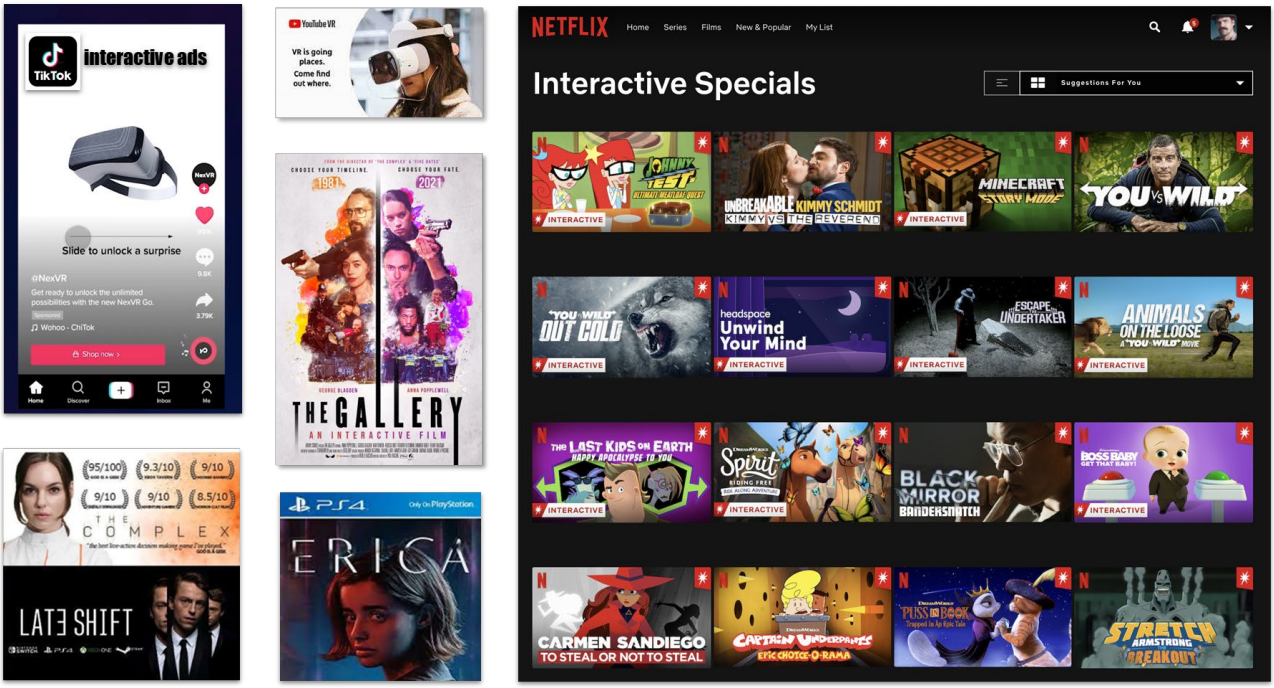
https://advanced-television.com/2022/10/17/peacock-unveils-interactive-storytelling-feature/

"The prototype combines the latest design vision with video samples from season two of the hit franchise. As viewers watch, Peacock will surface a wide variety of exclusive extra footage, timed to the most relevant moments, all with the user in control. As Peacock's first production partner, The Real Housewives Ultimate Girls Trip season three will launch with three episodes encompassing the final interactive feature. Users can lean back and watch an episode as-is or choose to engage with prompts as they appear at key moments during their viewing experience.

The new feature will first appear seamlessly in The Real Housewives Ultimate Girls Trip's main story stream through contextual opt-in dialogues. When the user selects a desired dialogue to further explore, they're taken into a side stream unlocking exclusive interviews, unreleased footage and never-before-seen cast reactions. As the viewer curates and completes their personalised experience, they're brought seamlessly back into the mainstream story."

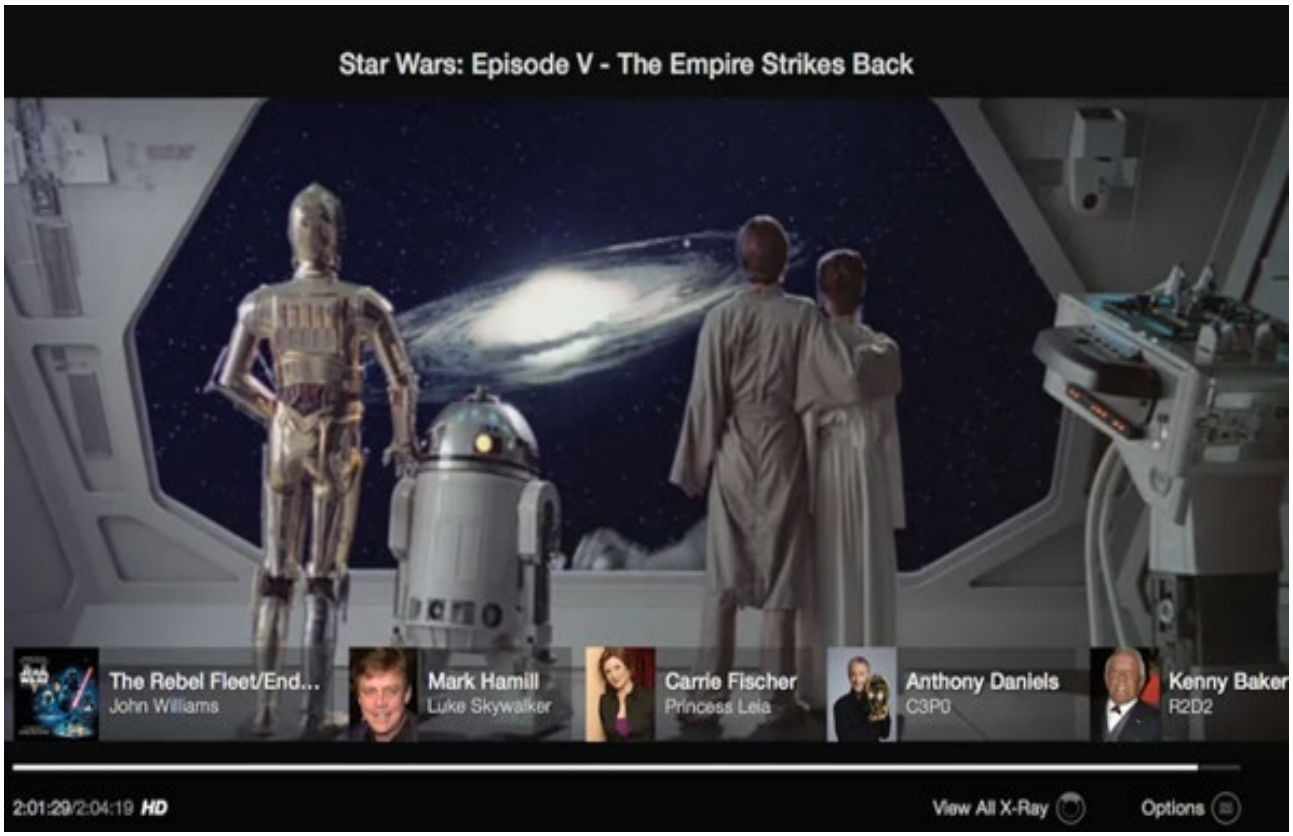
Another well-known example of this type of user experience, in a drama context, is Netflix's Black Mirror: Bandersnatch, where the storyline can go down multiple different paths depending on user choices. Other use cases include personalised news where the content could be tailored, e.g., to include relevant local items based on the user's location.

Netflix, YouTube, and others, as well as games engine tools already offer multiple path stories on their platform today. Much of the existing branching narrative content seems to be aimed at children as can be seen from the examples from the Netflix content gallery.



UC3: Associating additional elements/objects to a programme (Amazon X-Ray)

Amazon X-Ray provides an additional text and image layer for audiences, tailored to provide information about the on-screen subject matter. This can include character and actor information, cast and crew, information about the background music, or trivia about the episode or series. The information provided can change dynamically based on the context of the current time within the programme.



UC4: Distributing multiple versions of the same programme using OBM methods - changing some objects but not others (Swiss trial)

With the goal to optimise bandwidth use and enable personalised audio experiences in the future, Swiss broadcaster Schweizer Radio und Fernsehen (SRF) has been experimenting with the use of S-ADM in their SMPTE ST2110 based IP environment. A selection of open source prototype tools by Dolby were used to generate and analyse dynamic metadata in real-time. The balance of beds and objects of a typical sports production with multi-language commentaries were modified on a mixing desk, with changes travelling as metadata through the broadcaster‘s infrastructure.

https://tech.ebu.ch/files/live/sites/tech/files/shared/events/personalized_sound_experiences_workflows_workshop_2022/presentations/M_Brockmann_A_Weiss_2022_EBU_Event_Presentation.pdf

UC5: Live sports coverage of multiple events/venues with user interactivity (e.g., Olympics/Ateme)

This use case allows viewers to select their focus, e.g., when watching a multi-event tournament. For multi-sport events this would allow users to make selections of the particular sport that they want to watch closely, whilst potentially keeping an eye on other sports that are happening at the same time. In this use case the service would contain several video objects and audio objects corresponding to each sport which would be rendered at the device depending on the user’s selection.

HbbTV applications can already provide a solution for user interface aspects of this use-case. Such applications may have the capability to point the user to those different presentations, for rendering the corresponding media assets. This is however done in the context of an app, and the service from a DVB

perspective would then be defined either as an app only service, or as a service that comprises one of the possible presentations.

DVB systems are designed to cope with receivers with different capabilities, but it is important to note that displaying multiple video objects at the same time is not a trivial requirement in a receiver. The user interface shown in the use-case diagrams below imply that multiple video objects would be decoded and displayed simultaneously. This is one way to fulfil the use-case but other UI options may be available where only one video object is displayed at a time. For the purpose of this report we will focus on signalling of availability and corresponding properties of the media objects.

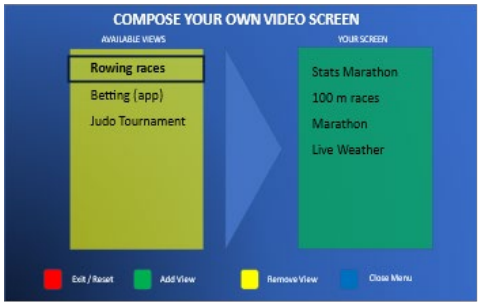
Web browsers may already allow for multiple video decode panels within a single window. However, while browsers on PC and mobile devices may be able to present multiple videos at the same time, this is often not the case on consumer televisions and STBs.

As part of the 2023 addition of the VVC video codec to the DVB specifications there is now support in the DVB codec toolbox (ETSI TS 101 154) for a mosaic of video images to be supported via a single video stream and a single video decode. It is therefore conceivable that the mosaic view shown in this use-case can be delivered using a single VVC stream for the video part.

To provide a consistent user experience across devices with varying levels of decode capabilities, cloud rendering could be used to offload the composition from the end user's device to the cloud or cloud edge. In effect, a single linear audio and video stream for each combination of elements to be displayed would be rendered in the cloud and distributed to client devices. The AI4ME project is investigating solutions in this area, to help efficiently deliver personalised streams in a scalable way.

As with almost all other use cases discussed here a default presentation should be signalled and delivered using existing signalling, ensuring that legacy devices will display the default media experience. It is expected that all the OBM functionality beyond the default presentation in the decoder would be optional.

For devices that support one or more additional OBM type functionalities, the viewer would be offered an opportunity to select other experiences besides the default.

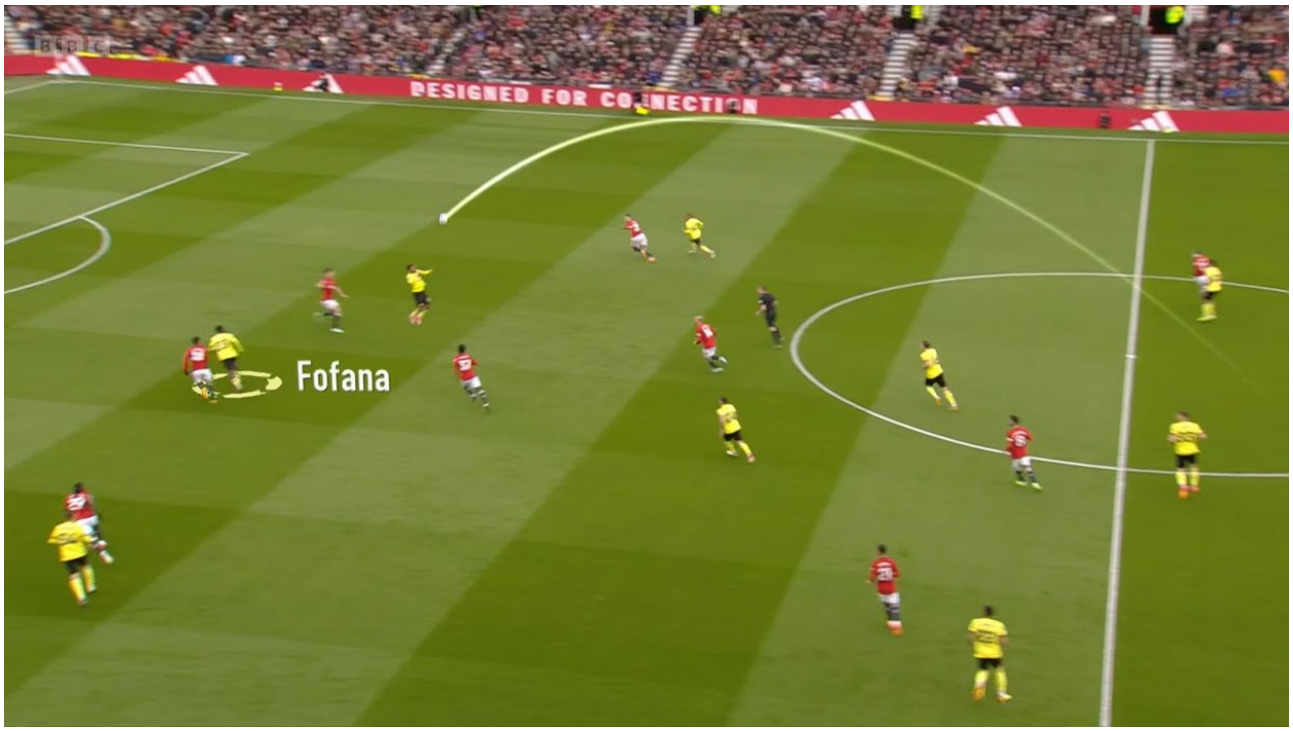


UC6: On-screen annotation of video scenes (Huawei)

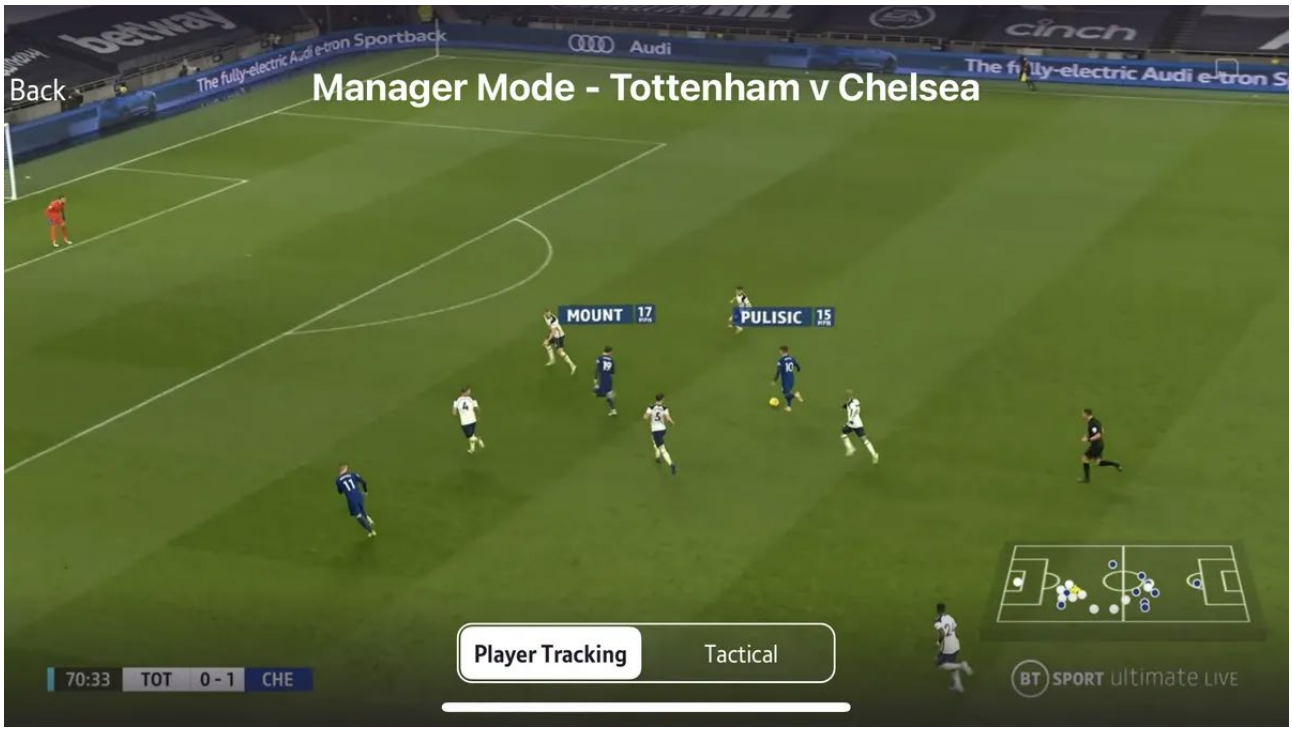
This use case shows the possibility of sending object layers which relate to the video being displayed. In the picture below it can be seen that graphics and data (Metadata) relating the performance of athletes in the video scene are being superimposed onto the underlying video.

Live athlete tracking mode - Image credit: Huawei

In these examples from Huawei, BT Sports, and BBC, the graphics are dynamic and show the name, position, speed, etc. of each athlete. Such information is now in common use for sports content, including the names and statistics of footballers displayed next to the player.



BBC Sports player tracking – image credit BBC Match of the Day)



BT Sports, player tracking mode selection - Image credit: Tom Wiggins

https://www.techradar.com/news/bt-sports-manager-mode-the-new-tv-tech-that-turns-football-fans-into-armchair-analysts

This kind of information has so far tended to be 'burnt-in' and rendered to be part of the video content in production, meaning that viewers have no choice to remove them.

With OBM, the information could be sent as metadata which would contain the information to be displayed, as well as the X,Y position on the screen of where the data should be displayed.

An alternative approach to carrying this data using OBM techniques would be to create a full page graphics layer with all the information presented in it, and then providing instructions to the receiver on how to superimpose this as a transparent graphics layer.

Web technology and bespoke applications could be leveraged to fulfil this requirement, but would not be suitable for applications where the content is distributed to multiple rights holders which may or may not want to offer the extended metadata to their viewers.

The utility of this use case is quite dependent on the distribution model of the content. For example the metadata object containing this information could be authored by the originator of the content (e.g. for the Olympics this might be OBS, or for the soccer World Cup could be FIFA) and would be carried alongside the main content. A broadcaster's app or video-player could then offer to viewers to enable/view this additional content or not. The ultimate choice of colours/fonts/size would be the choice of the decoder app, but the metadata would come from the stream.

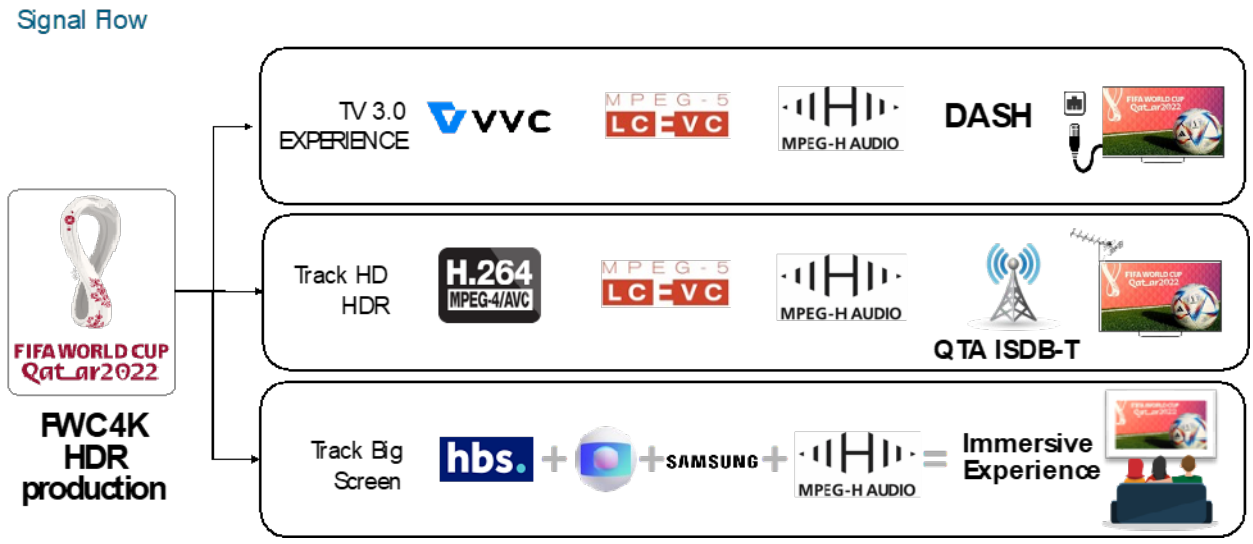
At its simplest this means that the object being delivered in this use case could merely be a set of metadata only, rather than any video or audio.

This would be similar to how information about which actors are in the scene are carried today by Amazon and others, but with the addition of screen reference (coordinate) data.

UC7: Next Generation Audio in action (UK, Brazil, Poland)

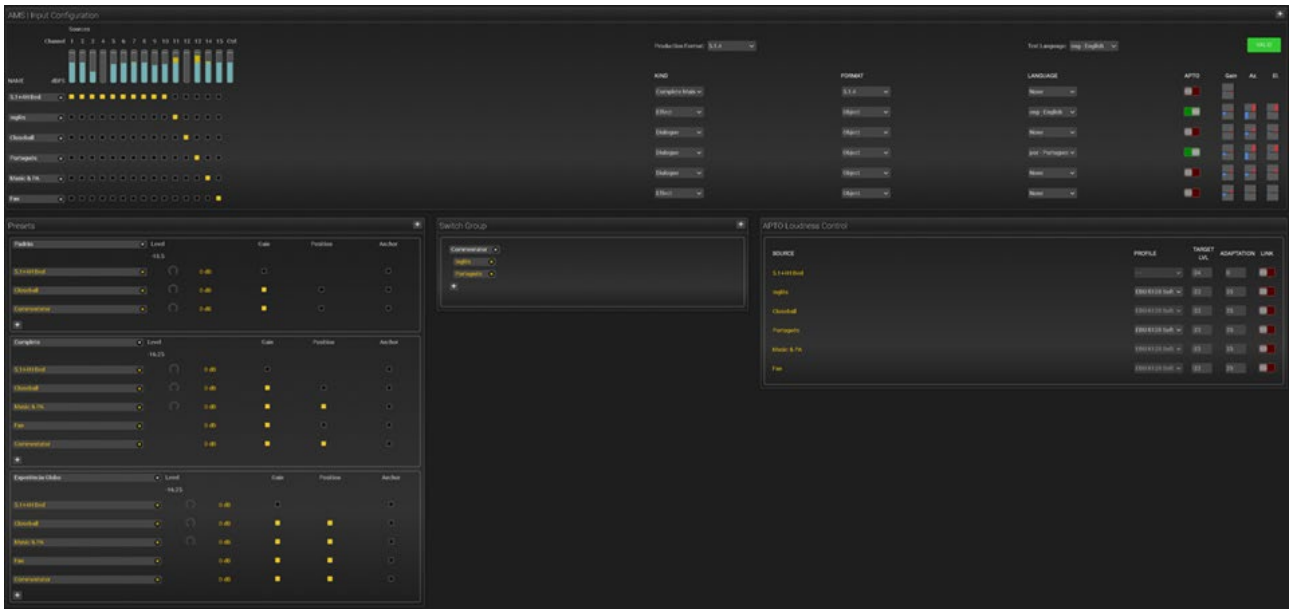
Next Generation Audio in Brazil during and since the 2022 World Cup.

Globo, the largest media group in Brazil, used MPEG-H Audio during the 2022 soccer world cup and since then continues to broadcast 24/7 with MPEG-H Audio. During the tournament, Globo integrated the MPEG-H production workflow into their Main Control Room (MCR) for coordination of three major distribution channels. A single MPEG-H production is used for distribution over the existing ISDB-Tb terrestrial broadcast system (TV 2.5) as well as DASH streaming, together with VVC video as the first TV 3.0 test (next generation broadcast system in Brazil).



Courtesy of Globo

The immersive mix (5.1+4H) together with additional audio objects for the Fan and Close ball sounds were received at Globo MCR room where the additional audio object containing the Portuguese commentator was added. Using the Authoring and Monitoring System (AMS) from Linear Acoustic, the MPEG-H Audio scene was authored to enable various personalisation options based on the available audio objects for each game.



MPEG-H Authoring and Monitoring system screen.

The MPEG-H authoring process included the creation of three audio pre-sets, as described below:

- Preset "*Padrão*": composed of an immersive 5.1+4 mix, an option to select the match commentator language (English or Portuguese), and an option to enhance the sound of the kickball (Close Ball object).
- Preset "*Completo*": identical to the previous preset, with the addition of audio objects with fans cheering sounds (Fans object) and pre/post-match music (PA music object). In this preset besides the level control enabled for all objects, the Fan and PA music objects were authored to allow the control of their position within the 3D space.
- Preset "*Experiência Globo*": identical to the previous preset. The authoring of this preset allowed the total position and volume control for all the audio objects in the audio scene.
- In addition, commentary was available in a choice of languages, Portuguese and English.

TV with integrated MPEG-H user interface.

Next Generation Audio in Poland for FIFA World Cup 2022

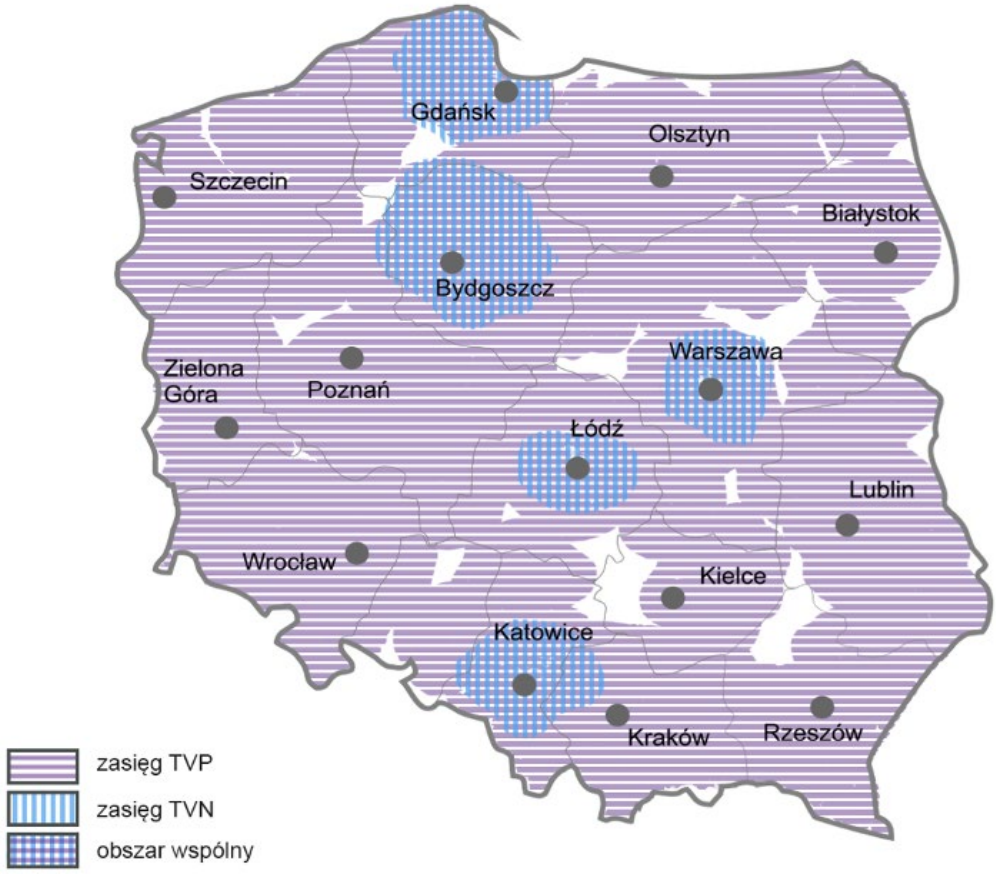
In 2022 the polish national broadcaster TVP delivered live sports in 4k video and next generation audio with Dolby AC-4 via DVB-T2.

The service contains the following components:

- STEREO (E-AC-3 Stereo)
  - Polish TV commentary
- DOLBY ATMOS (E-AC-3 Joint Object Coding 5.1.4 immersive)
  - Polish TV commentary
- DOLBY ATMOS (AC-4 NGA 5.1.4 immersive with personalized experiences via Preselection)

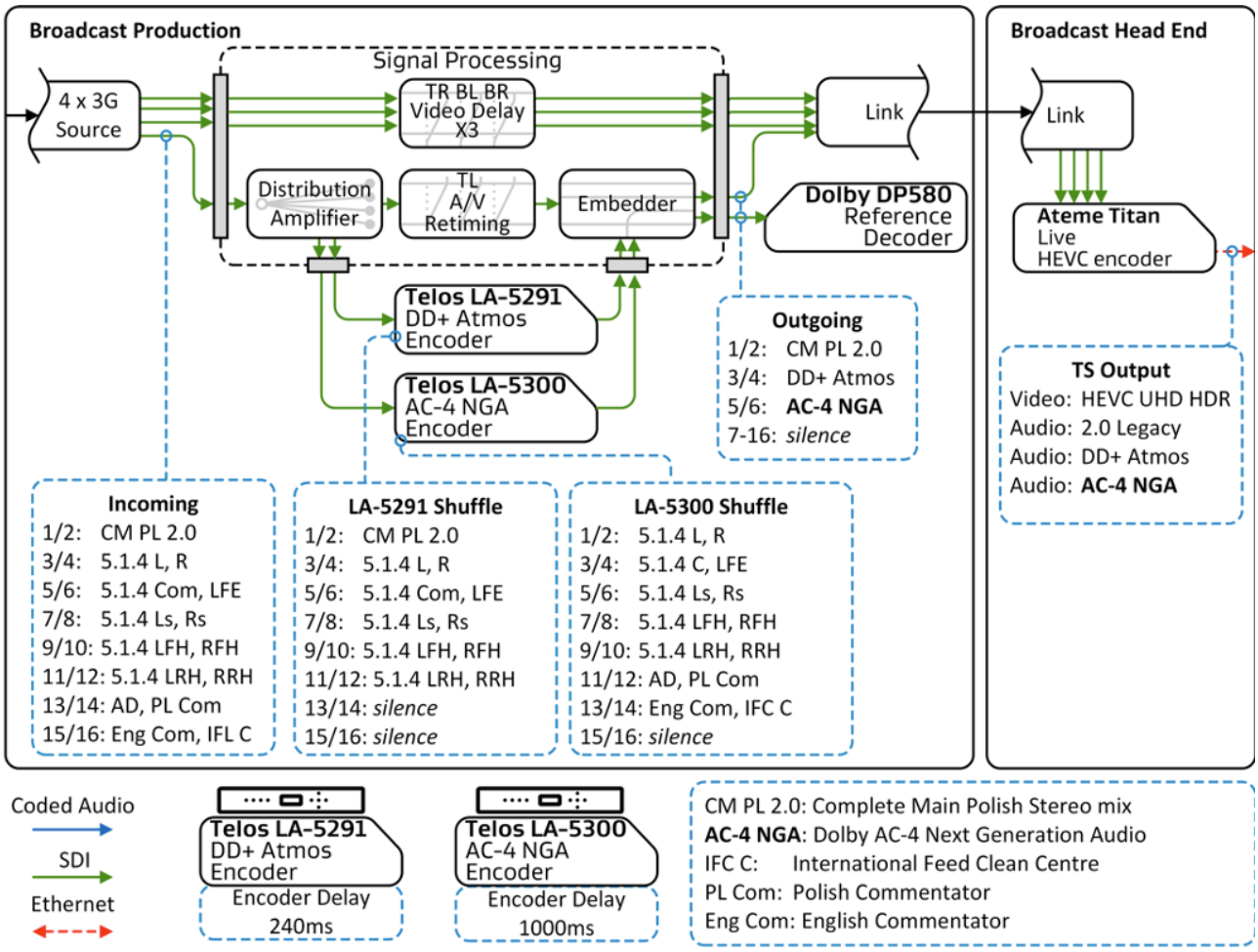
- Polish TV commentary
- Polish Audio Description
- English TV commentary
- Stadium only (IFL-C)

The nationwide availability of the TVP services are shown in the following map.



The nationwide availability of the TVP services – Image Credit TVP

The whole workflows were based on existing hardware as shown in the following illustration.

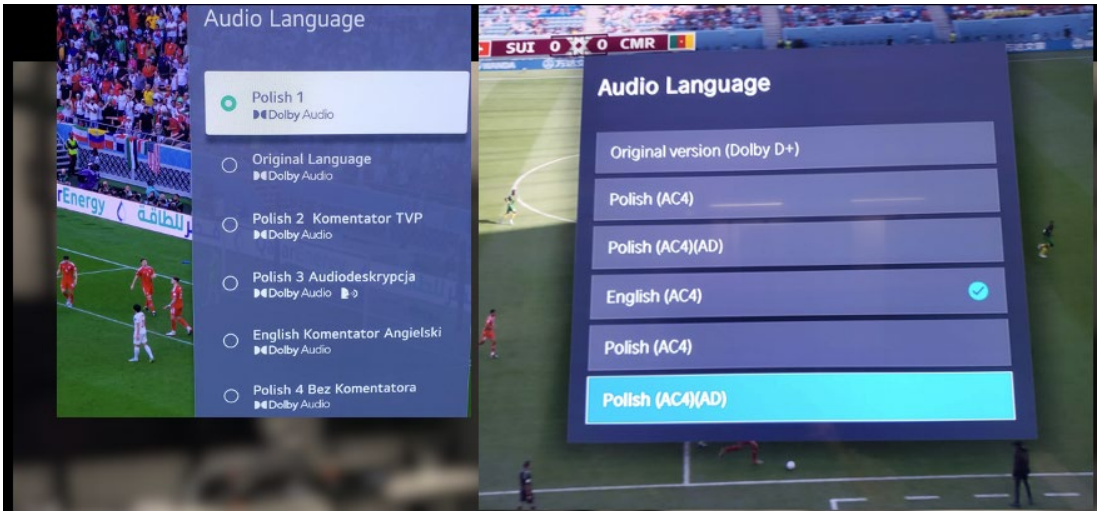


TVP NGA workflows were based on existing hardware - Image Credit Dolby

The DVB-T2 transmission included standardized DVB signalling for the NGA Preselections.

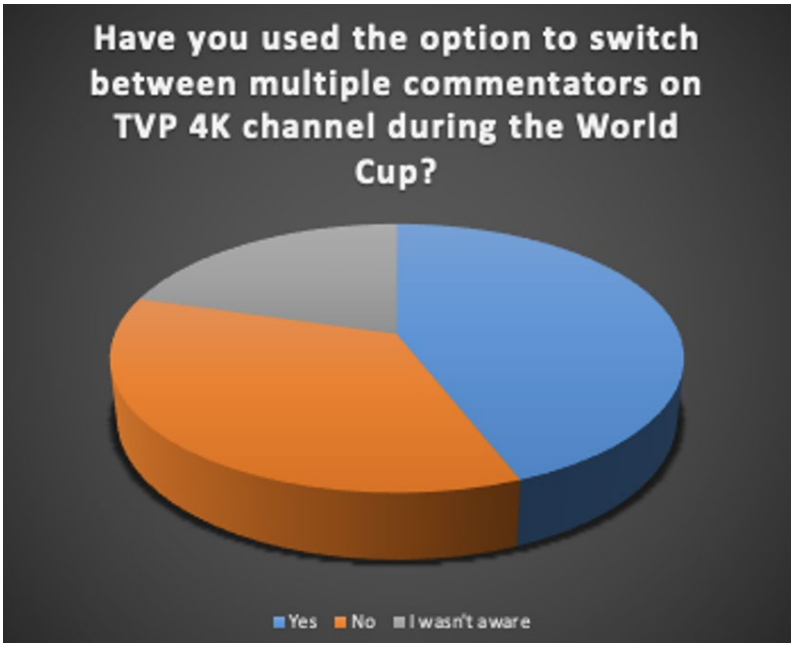
Audio Preselection Descriptor as defined in ETSI EN 300468.

TVs in the field support this signalling and provide user-interface for switching between personalized experiences.



User interfaces of Commercial TVs showing the TVP NGA service – Image credit Dolby

A TVP Survey showed that about half of the viewers have used this new functionality.



TVP Survey showing if viewers have engaged with NGA – Image Credit TVP

Next Generation Audio delivery to Second Screens

It is possible to deliver some audio content by broadcast and some by IP, and to keep the two synchronous. Use cases include delivery of additional languages or audio description where broadcast bandwidth is limited. The content can be reproduced on two separate devices, or within a single device.

It is possible using HbbTV for the receiver to decode both broadcast and IP delivered audio at the same time. This can be achieved by a MediaSynchroniser function in the HbbTV terminal. It is explained in Chapter 13.2.2 of ETSI TS 102 796 V1.7.1 (hbbtv.org). There are multiple options for transmitting the "master" timeline, but recent implementations have used MPEG-TS Timed External Media Information aka TEMI for a mixed DVB (broadband and broadcast) mode. Therefore, in our example the HbbTV terminal was receiving a timeline via TS and was requesting the right DASH-Stream and feeding it to the same decoder.

UC8: Interactive Sports matches – BBC iPlayer

In November 2022 the BBC started providing World Cup viewers with new Object Based Media features while watching live via BBC Sports Live webpages.

Features included an in-game stats layer, a player line-up view, chapters and key event markers to show information such as team possession, player selections, pivotal referee decisions and goals. Spoilers are taken into consideration, with markers only being shown on the timeline once the playhead has reached them.

BBC Sports Live webpages with Object Based Media features – Image Credit BBC

Over the duration of the 2022 tournament, over 950,000 unique users visited the live experience. Around 30% of those were repeat users on two or more games. Users rated the experience at 4.1/5 (from 16,000 user ratings).

User testing interviews were conducted over the full length of the tournament with participants from across a wide mix of geographic locations and were aged between 20 and 50.

Time spent with participants involved in-depth conversations about their current viewing behaviours, values, expectations of the new viewing experience, and use of multiple technologies while watching live events. This user research has been used to further develop potential solutions for chapterisation and key moments in other content genres.

Evidence showed that it brings real benefits to audiences. It can improve their enjoyment of the match, reduce pressure regarding fear of missing a goal when leaving the room, allow people to watch at their own pace, and help support sports-orientated conversation after the game.

These features are currently available on the web for desktop and mobile browsers. The BBC is currently working on how it would achieve this form of chapterisation and user-controlled information overlays consistently across web, mobile, and smart TVs in iPlayer and Sounds, and for different content genres. Ongoing work also includes looking at the data sources used to drive these content experiences, including any potential standards needs.

UC9: Using DASH with branching narratives and versioning

MPEG DASH is one of the most commonly used streaming protocols for media, but the DASH protocol, and DVB's own DVB DASH profile, has been primarily used for linear (non-branching) narratives.

Its largely experimental use for branching narratives has ranged from having one long timeline and jumping through this timeline based on user choices, or creating multiple full versions of the programme where the user choices trigger the device to stream an alternative (full) version from the DASH manifest.

The industry and particularly MPEG have recognised the importance of supporting branching narratives and versioning in its DASH specification. This has spawned new extensions and work at MPEG to cater for such a use case.

A description of the technology solution for branching narratives and versioning using DASH is described in section 6.5.3

UC10: Infuse Video & BBC & EZDRM Responsive Narrative Factory

At IBC 2023, Infuse Video (website) demonstrated a form of object-based media that enables personalized video experiences.

It was based on a challenge that originally had two objectives that were merged:

Personalized media/flexible media (allowing users to combine or operators to combine).

Filtering certain programming (e.g. nudity, violent content, content sensitive to users with epilepsy, age restriction).

A video of the discussion is available here: Accelerator Project: Responsive Narrative Factory - IBC2023

Topics covered include:

1. Generation and embedding of metadata to get all the tags that describe a piece of content.
2. A real-time online editing tool, used to create the targeted presentations with filtered or added content.
3. DRM, supported by EZDRM.
4. Rendering at distribution versus rendering at consumption.

The demo included a BBC Springwatch video that allowed users to select which birds they preferred watching. As a result the video only contains clips with these animals. Also the users can select short duration versus long duration clips. (See Fig. UC10a).

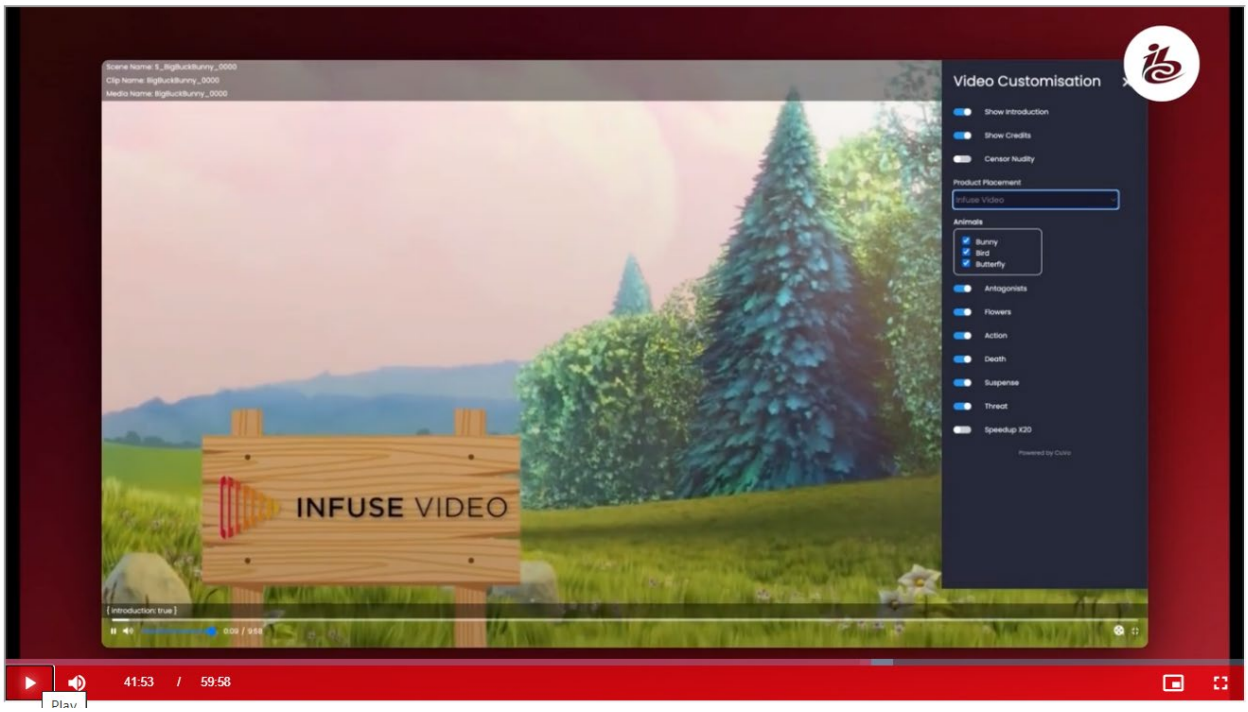
Another example usage includes personalized in-stream advertising, where specific scenes of the content are rendered differently to support a promotion or advertisement. In fig. uc10b you can see the popular movie Big Buck Bunny with an overlay advertisement.

A third use-case shows an example of personalized in-stream filtering. In this case specific scenes of the content are filtered out or changed. In fig. UC10c you can see the same movie without the actual bunny or text referring to the bunny, this content was filtered out. This is easily enabled in the demo by switching each element on or off. In addition other aspects can be removed such as occurrences of death, credits at the end, etc.

An important advantage of the solution is that the content is created on-the-fly in some cases or substituted in, avoiding the need to create different copies of the entire movie for each configuration. Thus, contrary to adding subtitles and/or graphics overlays, the customization of video potentially leads to large overheads in data and storage. The solution proposed by Infuse Video using online editing is one way to reduce some of this overhead. Overall in case flexible media / object based media is going to be considered, the storage/duplication part or requirement should be addressed as well.

Some consideration of the general media chain should be considered for this use case, such as for example relating to content rights and essence of content (i.e., can we just remove certain content) and also from the user experience, i.e., what are the best ways to improve the user experience for this. Overall, it seems that this technology can be used to generate more personalized and preference playlists based on user interest or configuration.

UC10a: Dynamic media annotated with metadata to enable selecting presenter, animal and sensitivity filters (by infuse video)



UC10b: Dynamic media where a scene is used for a localized promo or advertisement – Image Credit Infuse Video

UC10c: Dynamic object-based media where a scene is used to filter certain content, in this case the bunny – Image Credit Infuse Video

UC11: Social Media Share

Popular streaming platforms such as YouTube make it possible to share a link to a specific moment within a piece of content; this isn't currently implemented for content distributed by broadcasters. TV viewers cannot share a specific moment. Instead, they have to wait until the programme has finished and is made available on demand, and then share a link to the entire programme. This does not meet the viewers wish to immediately share a specific moment, the demand for which is illustrated by the way consumers take photos of the TV screen and share them. Implementing a "share" function would drive traffic to broadcasters' content and provide useful data on the sharing that happens. Media objects can include segments with varying degrees of granularity, making it possible for broadcasters to introduce "share" functionality on many platforms.

8. Gap analysis

When considering what is possible with existing DVB standards, this study mission broadly categorised media types as either Foundational or Extended.

- **Foundational media types** should be seen as the core essences of the content experience. Traditionally these will be the video and associated main audio mix. Main language accessibility, Subtitles and Audio Described (Narrative) track, are now also considered as foundational media types by virtue of their prevalence, standardisation, and core need in society today. These foundational media types are necessary for the basic tailored consumer programme experience (e.g., typical linear broadcast programme or in many cases for a comparable streaming experience). They already have well defined and widely deployed delivery mechanisms within DVB specifications for both broadcast and IP delivered content. In Object Based Media, these foundational media types would be delivered using the existing delivery specifications, no new specifications are needed. When foundational media types are delivered as part of an object-based media presentation, additional metadata may be delivered to explain how they can be replayed and combined for an object based experience. They may also have additional time-based content appended, added or inserted into the timeline to allow for personalisation based on user interests, creative intent or territorial needs (e.g., titles or credits in a differing language, scenes amended due to age or territorial regulations, or to enable interactive selection of additional content).
- **Extended media types** may be considered as those that do not necessarily have existing DVB specifications to rely on. Thus where used today, the implementations may be considered proprietary. Extended media types may be media data which consist of information such as Haptics tracks to drive vibration and haptic feedback devices if and when they exist in e.g. mobile devices or dedicated devices, Transmedia Objects such as actor profiles, synopses, trailers, or alternate experiences (be they audible tracks, branching narrative structures, or graphics) selected by the user to ultimately construct their desired experience.

8.1. What is possible with today's standards

Standardised point solutions exist for many types of objects today. These include audio description, subtitles, targeted advertising, now and next information and EPG data.

Standardised solutions for Next Generation Audio are also maturing, although the handling of NGA in distribution is somewhat specific to the audio codec being used.

Companies have also taken standardised solutions such as DASH and adapted these to carry multiple objects.

Web standards offer some useful functionality as do some app frameworks such as HbbTV, so far however developers seem to be relying on a combination of frameworks that they have developed for use within their own applications, which do not rely on any standard, and where interoperability with other applications is not a consideration.

8.1.1. Gaps shortfalls and missing functionality today

The existing DVB specifications for signalling of the foundational media types (e.g., Subtitles, Signer or Audio Description) make a considerable amount of user selections and options possible today. All of these existing solutions work well but are independent of each other, they nevertheless can be combined in a single DVB DASH manifest or DVB Transport Stream.

For audio, the DVB specifications may not have significant gaps, because with the introduction of Next Generation Audio, corresponding object based use-cases were added. Gaps were identified for video, when it comes to multiple video feeds to allow enhanced personalization of the video experience. For Extended

Media Types that go beyond the foundational ones (audio, video, subtitles) the DVB specifications don't provide corresponding signalling today.

Foundational media types are already delivered in existing standardised ways, but today the methods used identify them and link them to the primary programme or to each other often rely on proprietary or app specific data.

At this stage no commercial requirements for an OBM specification have been received, and some members of the study mission question whether additional specifications would be needed, others consider such an OBM specification would help the industry.

The AI4ME project is investigating compute-aware content delivery as a way to efficiently deliver personalised object based media streams to large audiences. Computation workloads, such as the dynamic composition and rendering of object based media content can adapt to the availability of compute capability, e.g., in consumer devices or in the cloud or edge. This could point to a potential need for standards for device capability detection, negotiation between devices and networked compute resources, media delivery, and low latency interaction. The project presented their use cases at the IETF Computing-Aware Traffic Steering Working Group meeting in November 2023.

8.2. Outlook horizon use cases for OBM in 3-5 years from now using forthcoming standards

It is clear that further use cases for OBM are still surfacing, and through a combination of apps and proprietary platforms, are being fulfilled, especially for online delivered content. The next three years look promising for additional use cases using the now established and well-deployed NGA frameworks, and it is expected that genres such as sports will continue to embrace the possibilities these technologies bring. Further use of games engines in the delivery of entertainment content may also expand the possibilities for OBM.

Although some standards are under development, given the current trajectory it is unlikely that these will make a major impact as they are only aiming to solve some immediate gaps and are not planning to provide a holistic solution.

In the absence of this we expect that web apps, browser based technologies and dynamically installable apps will be leveraged to fulfil some of the emerging use-cases.

8.2.1. Gaps shortfalls and missing functionality

From a DVB perspective, focusing on content discovery, delivery, and media decoding, the biggest gap is currently the absence of a universal mechanism to associate different media elements together, and to provide metadata to identify them and indicate if and how those elements can be used in playback. This mechanism is a prerequisite for any further developments such as synchronisation and personalisation. In other words, without a list of the items available for a given program, very little innovation in the manipulation, personalisation, sync and playback of those elements is possible.

9. Survey

During Q1 and Q2 2023, the study group ran a survey on Object-Based Media to gauge the industry's interest in the different categories of objects discussed by the group and to gauge the respondents' attitude towards any potential standardization effort in this domain.

In total there were 16 questions with 4 of these questions for admin purposes and to qualify and categorize respondents. The survey received 14 responses.

9.1. Introduction to the survey

DVB survey on Object-Based Media

The DVB Project is undertaking a study mission to gain insight and to collect use cases for the uses and applications of Object-Based Media (OBM), a term that covers the delivery of multiple elements (objects) that are related to the same programme/content. In DVB systems today there are specific solutions and specifications for carrying the most common objects, but these are only designed for that single, specific purpose (such as carrying subtitles) and are not normally used for other purposes.

Some objects may be seen as the default content, such as the standard audio or video, while other individual objects might be considered to be additions, such as accessibility elements like audio description/signing/subtitles or an alternative language.

Further additional elements that can be added to the playback and that are related to the service (cast and crew information, synopsis, still image or video trailer) can be defined and treated as objects. Today, most programmes contain at least two additional objects (e.g. audio description, subtitles) but many already contain several more (such as visual signing, director's commentary, additional graphics, EPG links, trailers and references to online information).

The DVB OBM study mission is examining how objects are being used in the industry and determining:

a) how existing specifications cover the use cases for carrying programming with multiple objects, and

b) whether there are commercial drivers that will justify further standardization by DVB.

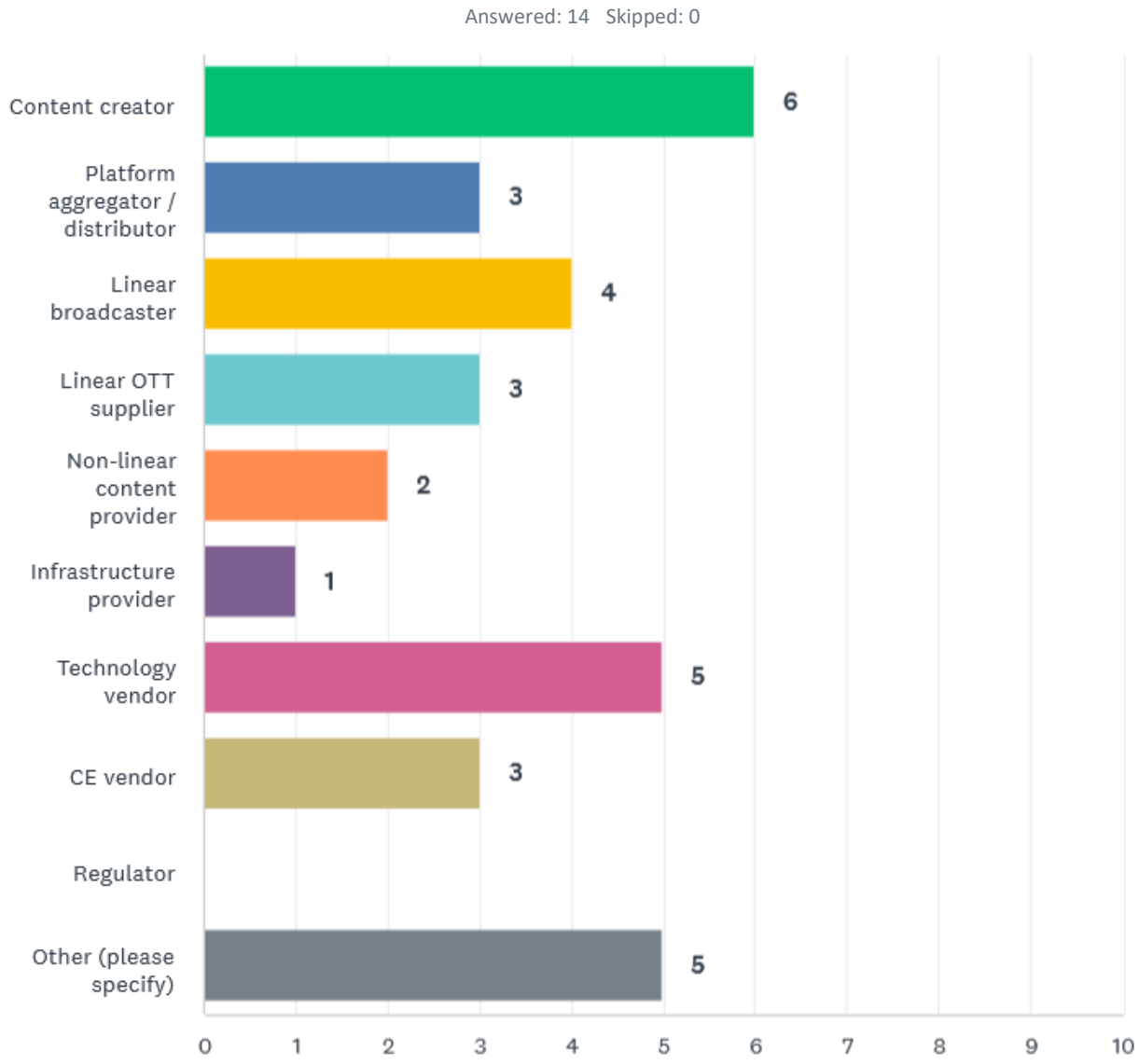
The following questionnaire seeks to poll respondents to understand which use cases for OBM are in use today and which may soon need to be supported.

9.2. Responses

The 14 responses received were from a wide cross-section of the industry with at least one respondent in every industry category offered.

Q3

What does your organization do? (check all that apply)

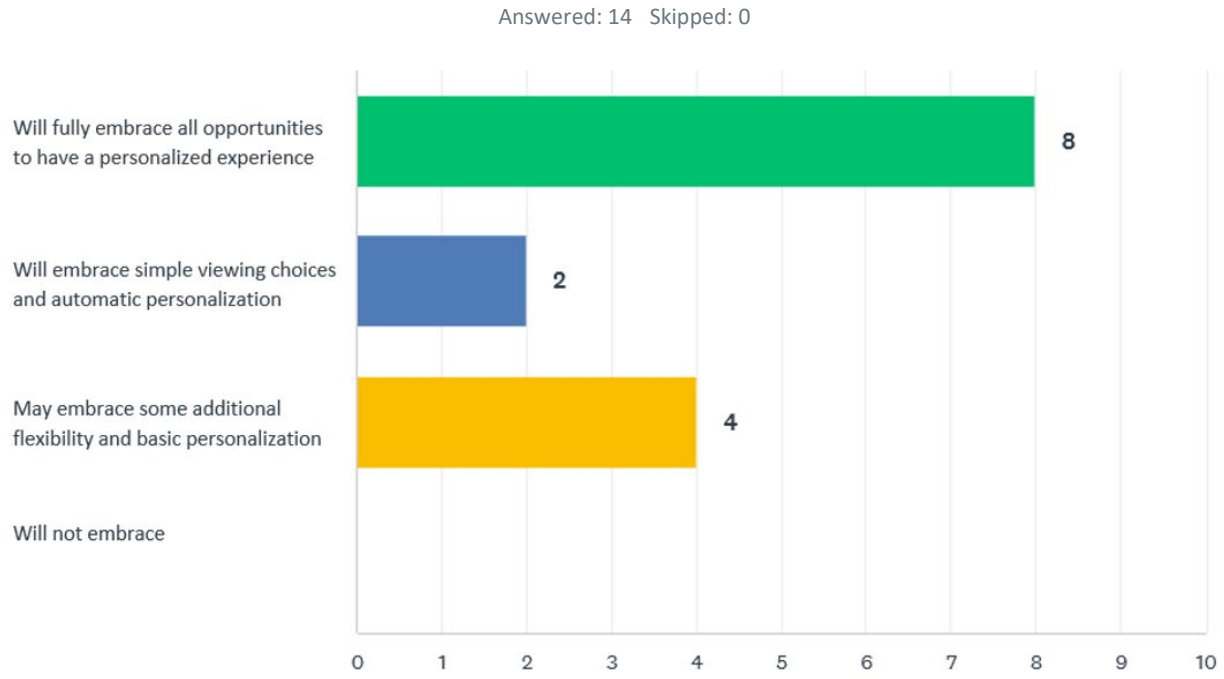


Study Mission interpretation Q3

The survey has provided responses from a broad cross-section of the media delivery ecosystem, and although the number of respondents were small, the representation is broad. The content creator category is particularly well represented which is reassuring, as OBM content has to originate from the creative process.

Q4

Consumer adoption: Considering that OBM allows the user to have a tailored experience (in an active manner by directly making selections, or passively by allowing selections to be made based on known preferences), how do you see consumers adopting the concept of tailoring media consumption for their own media experience rather than just have the default creatively constructed programme?

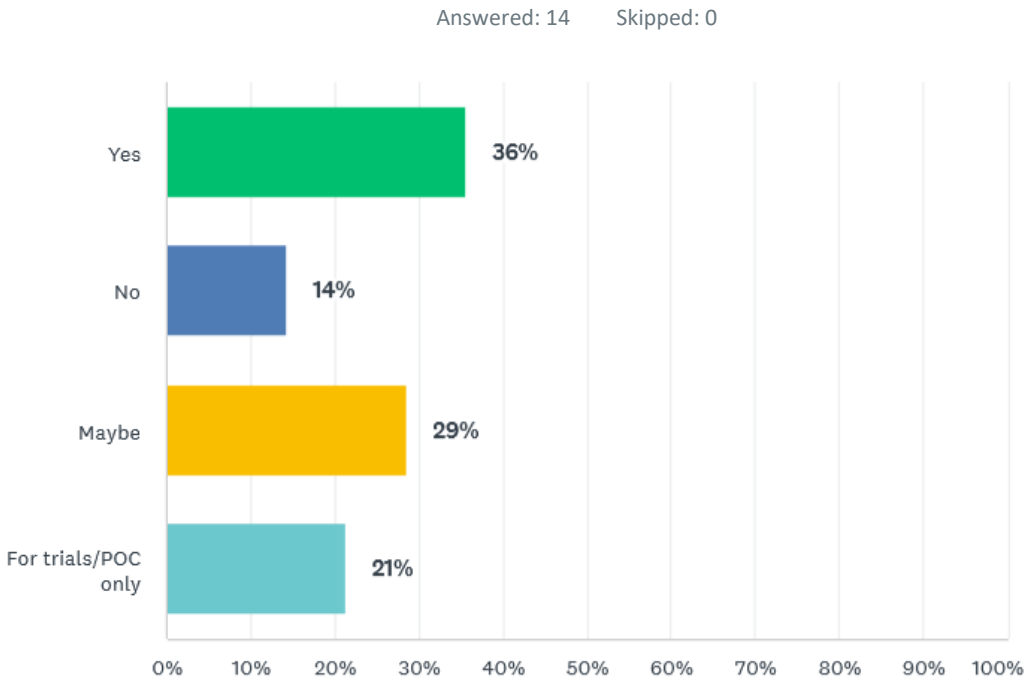


Study Mission interpretation Q4

None of the respondents predicted that users would not embrace the possibility to personalise their experience. We can therefore assume that survey participants expect the provision of personalised content to increase. The survey shows that over 50% of respondents expect users to fully embrace personalization whilst others feel that users are interested in basic choices and customizations. In practice the level of personalisation is likely to be dependent on genre, and perhaps the device on which the content is consumed.

Q5

Branching narratives: Some established OTT platforms support and offer limited programming with branching narratives (where the user is provided with options to navigate through the content e.g. news stories, or alternative endings). Does your organization consider programming with branching narratives (alternative paths) as being important to the future content delivery?

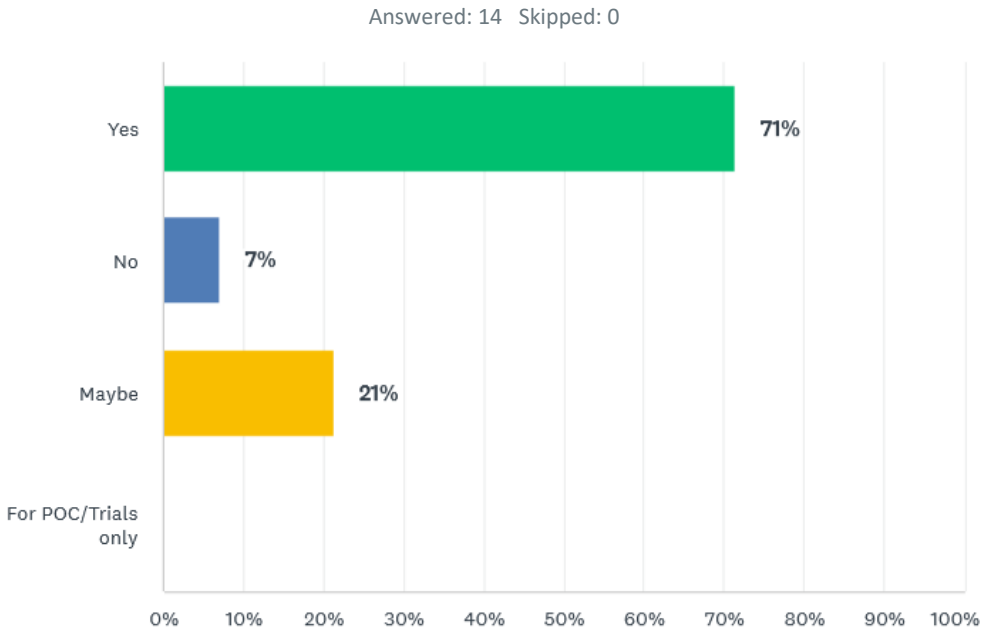


Study Mission interpretation Q5

Although 35% of respondents feel that branching narratives will be important to the future of content delivery, the other two thirds are less convinced or less certain. We can therefore conclude that branching narratives is unlikely to become ubiquitous in the near future, and that perhaps it will remain a tool used for certain genres, or applications. We have seen many examples of children's content, educational content and the navigation of news/sport content. The survey does however show that only 14% of respondents feel that it is not of any importance.

Q6

Additional or alternative layers: Some platforms can support and offer some programming with additional or alternative layers such as optional graphics or additional audio tracks (where the user is provided with options to select their most favoured, tailored experience). Does your organization consider programming with multiple elements/objects as being important to the future of content delivery (e.g. alternative or additional languages/commentators/graphics layers/audio tracks/camera angles?

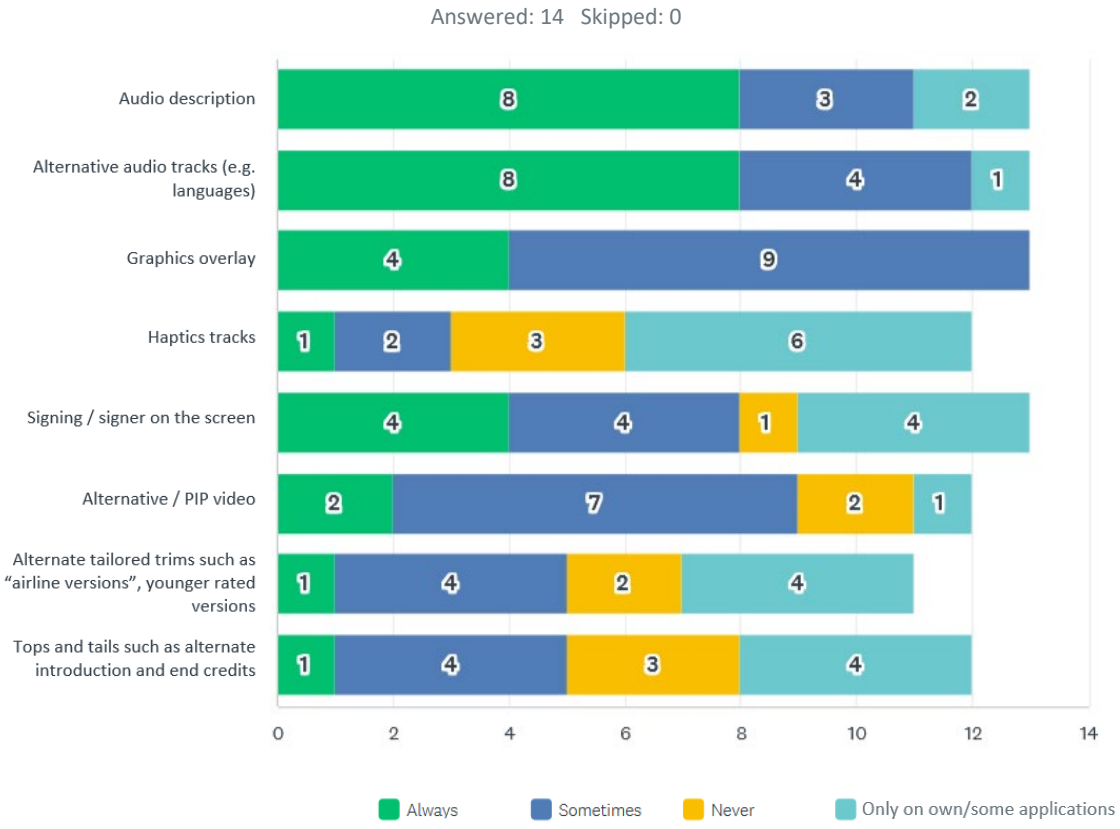


Study Mission interpretation Q6

71% of respondents considered the use of layered objects to be important to the future of content delivery. Although this is very impressive it is reasonable to conclude that this may be due the fact that subtitles, audio description, languages, and targeted advertising, delivered as separate elements which can be construed as objects, are already in widespread use. The question did however highlight objects that are less used today, so it is reasonable to assume that the range of objects and therefore playback experiences or options offered by content providers is set to further grow. What this question does not answer is how many of the respondents feel satisfied with their ability to deliver layered objects and using which platforms or mechanisms.

Q7

Support for layered elements: For programming today, in addition to the base elements of main audio and video, which of the following layered elements do you support the delivery of, or will you wish to support?



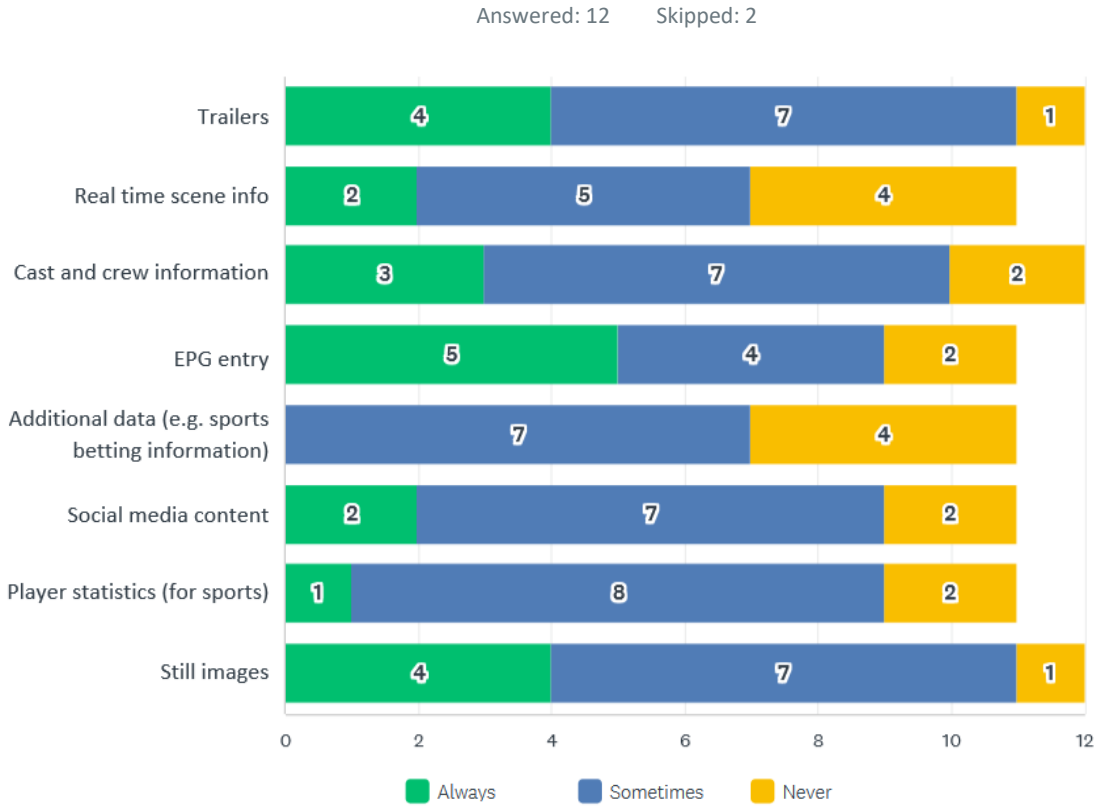
Study Mission interpretation Q7

It is not surprising that the layered objects seen as always being the most important in this question are those related to accessibility, and audience reach. It is however interesting to note that graphics overlay and picture-in-picture score highly in the "sometimes" category. This perhaps highlights the fact that such features are highly genre dependent, and might be very useful for sports or news, but less useful or even discouraged for movies and drama.

Newer media types such as Haptics seem to score well in the "only in applications" category, perhaps pointing to some functionality being reserved for operator applications where perhaps the operator wants to provide an enhanced experience, or encourage users to stay within their app.

Q8

Additional elements: Many OTT platforms today provide methods to find out more about the content, e.g. which characters or actors are in the scene, additional player stats etc. For programming today, which of the following additional elements do you commonly support or deliver?

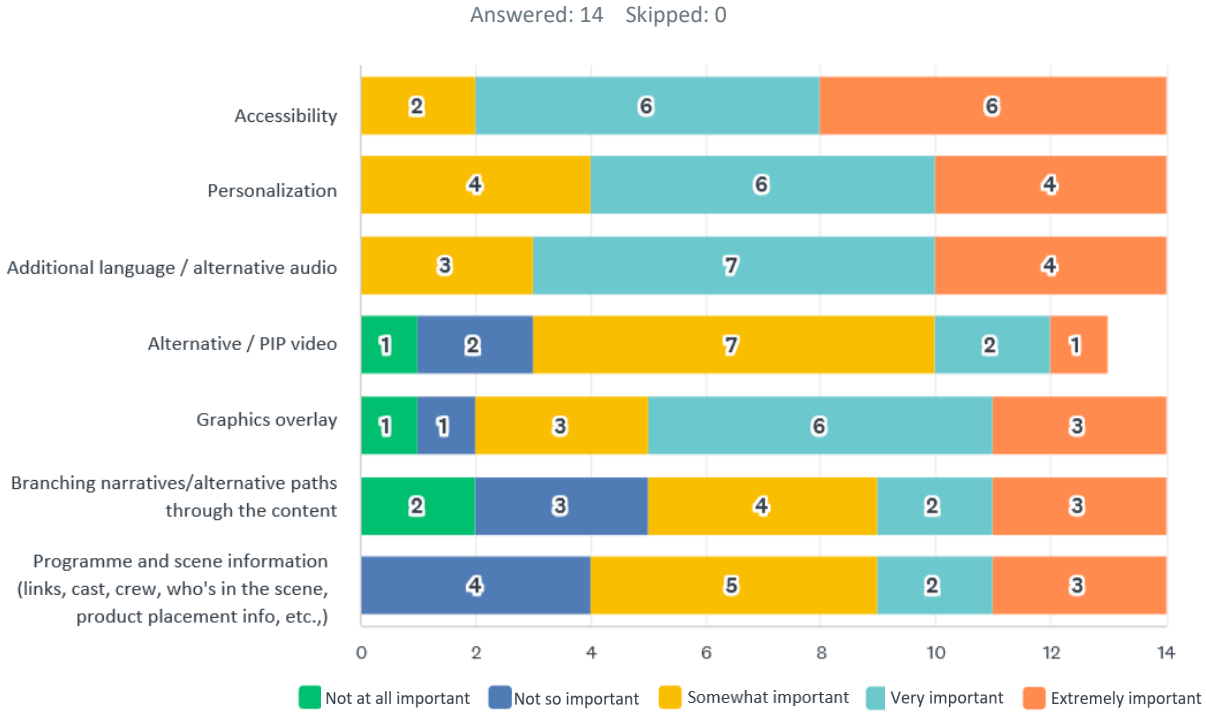


Study Mission interpretation Q8

The vast majority of respondents seem to already be creating/supporting or delivering most of the transmedia objects listed in the survey (we see this by adding the 'always' and the 'sometimes' categories). This means that the production of content that can be delivered as objects is already in place, and that delivering OBM is no longer a burden for production. What the survey fails to show is how the various items are stored and distributed today. We can assume that the social media content for example is distributed via the common social media platforms and not delivered as part of the program. It would be interesting to gain information on both how these content items are stored and to which platforms they are and are not shared.

Q9

Benefits of OBM: How important do you consider OBM in bringing benefits to viewers?



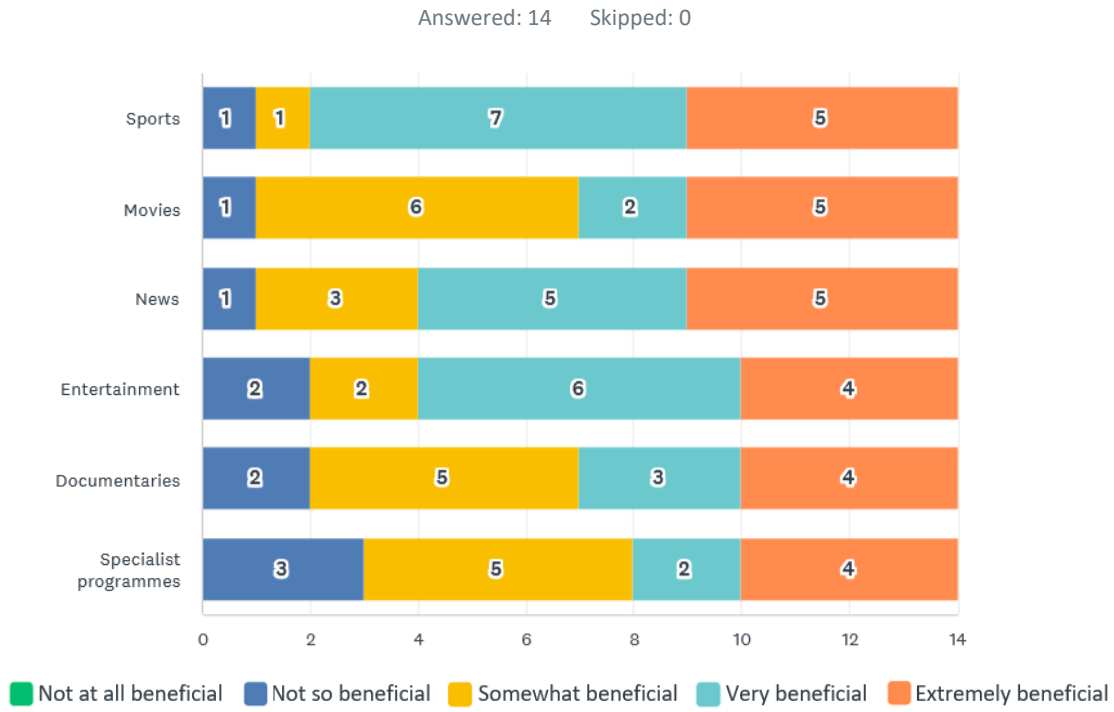
Study Mission interpretation Q9

All responses consider that the features listed are somewhat, very or extremely important. Some, e.g. PiP & branching are less positive but still >50%.

From this, OBM could be inferred as a good way to deliver accessibility services including extra language/audio and personalisation although some of the features can already be delivered today by existing DVB signalling so it is unclear from the survey if there are any 'additional' expectations.

Q10

Commercial benefits by genre: How commercially beneficial do you consider OBM could be to each of these types of content?



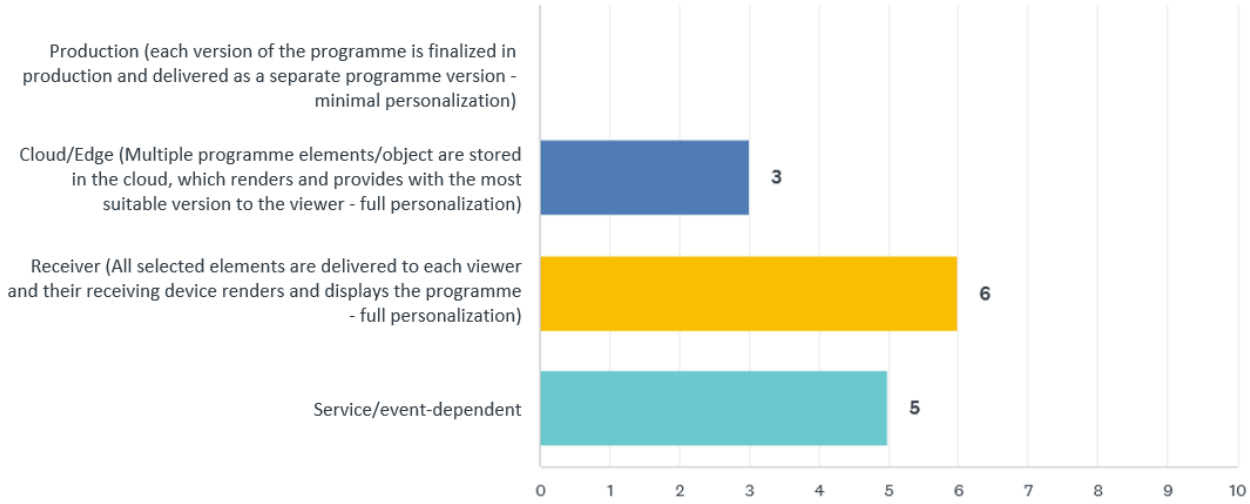
Study Mission interpretation Q10

The survey results clearly show that there is a strong interest in using OBM for virtually every category of content. Sports was overall the highest rated, followed by news and entertainment. The precise requirement on DVB is hard to perceive if it is accepted that foundational media types can be delivered using existing signalling mechanisms. The need for alternative or overarching signalling was not questioned in the survey.

Q11

Object rendering: Thinking of the commercial benefits to your organization in terms of quality of service to viewers, where should objects ideally be rendered?

Answered: 14 Skipped: 0



Study Mission interpretation Q11

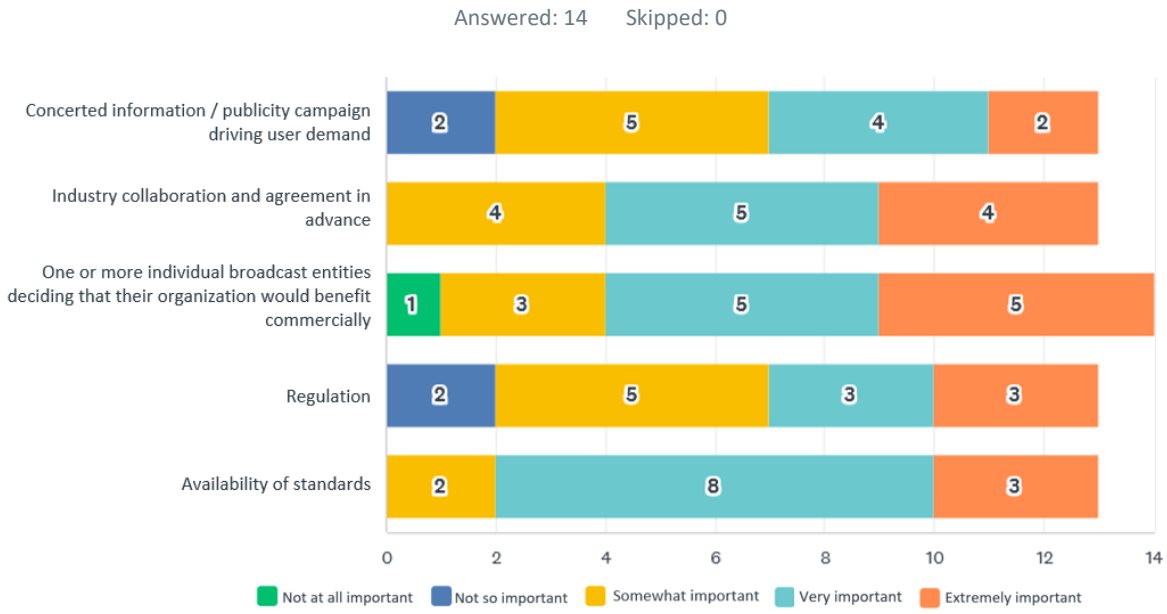
A significant proportion of respondents favour receiver side rendering. If this is related to foundational media types then this is already possible today. The last answer “Service/event dependent” essentially seems to require all three options which might be considered as ambitious.

It is noted that there was no demand for production side rendering only with multiple program versions then needing to be delivered.

With a 2:1 majority of respondents feeling it is best done on device compared to cloud /edge this should be considered as to the nature of the rendering as it may add some burden to typical/common entry level receivers which may not be accepted by all sides of the value chain.

Q12

Industry investment in OBM: Recognizing that an industry shift to OBM would require a significant amount of investment in upgrading, modifying and reconfiguration to broadcast systems and workflows, how important do you think these different drivers would be?



Study Mission interpretation Q12

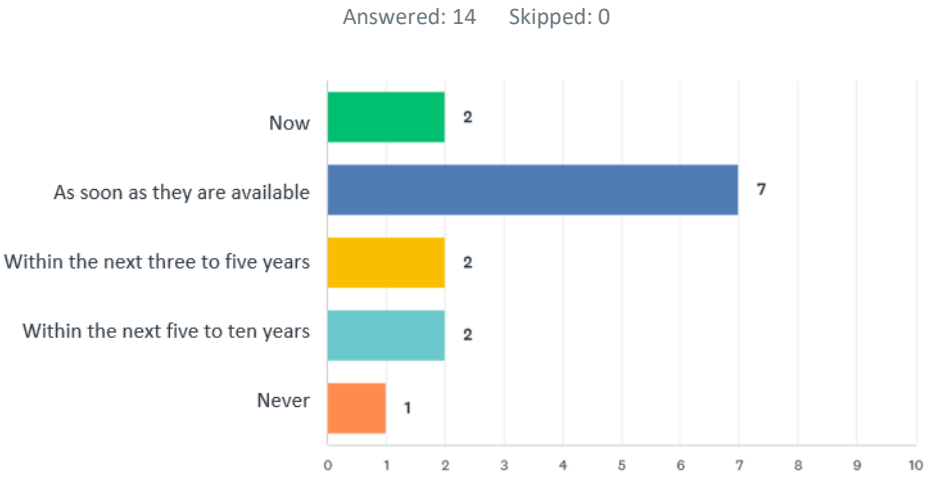
It appears from the survey that the consensus is that a complete cross-industry promotion plan would be needed, from specifications to regulation. It is hard to understand from the survey what the respondents would like to see in terms of regulation.

It is clear that all but one of the respondents felt that one or more individual broadcast entities deciding that their organisation would benefit commercially would be a major catalyst.

Should any work be contemplated on OBM, respondents would rank industry standards as an important concern, as well as industry collaboration in advance.

Q13

Benefits to your organisation - OTT/IP: If standards for the delivery of OBM (via OTT) were to become available within the next one to two years, when would these be of most commercial interest and benefit to your organisation?



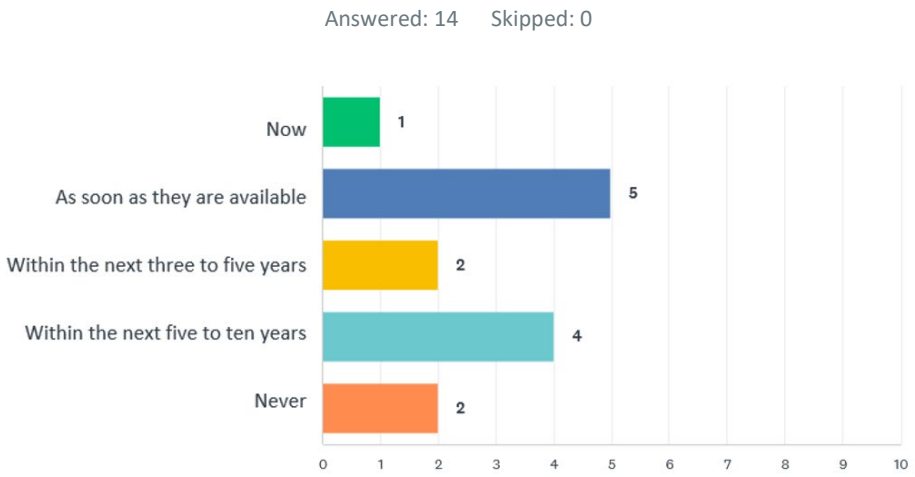
Study Mission interpretation Q13

The desire from respondents for standards is clearly very positive, almost 80% see the benefit from OBM standards in next 0-5yrs.

The exact use case of the respondents is unclear given that some OBM (e.g. subtitles) is already delivered today and has standards widely available. Without that clarification it is not easy to interpret the responses in a meaningful manner.

Q14

Benefits to your organization - Traditional Broadcast/TS: If standards for the delivery of OBM for broadcast were to become available within the next one to two years, when would these be of most commercial interest and benefit to your organization?



Study Mission interpretation Q14

There is reported to be a strong desire to use OBM for broadcast as soon as they are available (>35%) although the interest is notably lower than for the OTT option.

As with Q13, the exact use case of the respondents is unclear given that some elements of media (e.g. subtitles) are often already delivered separate from the media and can optionally be selected today and have DVB standards widely available. Without that clarification it is not easy to interpret the responses in a meaningful manner. It seems from the responses that most respondents are comfortable with a 3-5 year timescale, assuming that the standardisation process would take at least that long.

9.3. Summary results and conclusions (of the survey)

On the surface the results of the survey point to broad support for the concept of OBM ecosystems. The answers to many of the questions also hint at a need for some kind of standardisation. However it's unclear exactly what the respondents would like to happen first, and where the gaps might be. Given the broad range of respondents it is likely that many are already experimenting, trailing or have launched services incorporating one or more aspects of OBM.

Another concern is that the respondents may have a different interpretation of what is meant by OBM. Even within our study mission members, there was considerable discussion on what exactly constitutes an object based experience, and whether things we do today such as subtitles and audio description already count as OBM.

In terms of Genres (Question 10) where OBM was considered useful, Sports was overall the highest rated, followed by news and entertainment. This is borne out in the use cases that we received, where sports and entertainment feature heavily.

It was clear from Question 8, that most respondents were already producing many of the content and elements that could be used in OBM presentations, but this question gave little insight as to if and how these elements were distributed today, and on which platforms. We suspect that most respondents have a variety of delivery strategies across Web, Broadcast, OTT, and social media, but that all the media elements they produce for a program are rarely available on every platform.

In Question 9 Accessibility and Languages have clearly been identified by the survey as the most valuable aspects of OBM, but it is not clear whether the respondents considered the current widespread support for accessibility as already being Object-based, or whether they considered that more work was needed to elevate the accessibility further, or to improve the user experience.

10. Other commercial considerations of OBM

10.1. Content rights

OBM may have a profound effect on how the content rights are negotiated and licensed. For sports for example, there may need to be rights packages which include individual microphone feeds, as well as multiple commentators, languages, camera angles and statistics. This may need changes to the way content owners segment their content, and specify playback constraints.

For other genres, such as movie and high value drama having modifications or changes to the presentation have usually been restricted. To enable OBM with personalisation some flexibility would be needed. Content rights are not in scope for this study mission but should be considered, as a factor which may need to be resolved to allow OBM to scale.

10.2. Content protection/CA

The main programme video and audio content are often protected to prevent copying. For use cases involving branching narratives or dynamic composition of multiple video streams, there may be a need to ensure that all such Essence media objects can be protected to the same level. This may also apply to audio and video Transmedia Objects, but less so where such objects are text or images which have arguably less commercial value in and of themselves. There may also be cases where a combination of protected and non-protected content may need to be combined, or even content protected using different rights or a different system.

The group identified this as a potentially complex area for consideration, but did not undertake a more detailed analysis.

10.3. Archiving, & Monitoring

As Object-based media contains multiple pieces of media essence and many combinations of these, it is unlikely to be archived in the same way as simple single audio/video content, and should be archived as individual objects (including all metadata). Each rendered user experience can be archived using pre-sets, i.e., metadata describing how to combine those objects for a specific experience.

Because an object-based media delivery provides a personalisable experience, it is a much higher effort to monitor and QA all possible combinations and settings of a flexible user experience. However, when pre-sets are provided those can be easily monitored and QA'ed.

10.4. Regulation

The issue of regulation for OBM is beyond the scope of this report but some aspects were considered in the Ofcom 2021 OBM report (https://www.ofcom.org.uk/__data/assets/pdf_file/0026/223973/Object-based-media-report.pdf).

11. Concluding Remarks

The work of the OBM group is composed of 2 phases, the first of which is this published report.

In the second phase the group will prepare conclusions, and recommendations for the DVB project, some of which may be internal to the project.

Should you be interested in providing feedback on the report, please contact DVB DVB@DVB.org and the OBM group chair (OBM group page). DVB members can provide input directly to the group via its meetings.

12. History

Version	Date	[Milestone]
S102	May 2025	S102 BlueBook (internal document CM2354r2)