Ofcom
making communications work for everyone

Object-based media report

by Macroblock Ltd. and the School of Digital Arts

Published 6 September 2021

Macroblock

SODA SCHOOL OF DIGITAL ARTS

Manchester Metropolitan University

Contents

1. Purpose of the report

The report will look to the near-term future and attempt to provide insight into likely trends and technological advances that provide us with a view on the further likely development of Object-based media, together with a set of recommendations and actions that Ofcom may wish to take in order to ensure British viewers and listeners can fully benefit from these advances but remain protected as the technology rolls out.

1. To identify the current industry trends in object-based technology and the market for these.
2. To highlight what experiences and advantages these trends and technologies offer to the UK consumer.
3. To investigate the challenges and opportunities for the broadcast and communications industry in delivering the new experiences to consumers.
4. To outline any planning or provisions that are anticipated for major world events (e.g. the Olympics) that may act as a catalyst for the use of these technologies.
5. To highlight any impacts or actions that are needed from a regulatory perspective with regards to consumers and the industry.
6. To provide a set of recommendations to Ofcom and the UK media industry on steps to take and further work needed to ensure that the UK is well placed and ready from a regulatory point of view to take advantage of the advances in Object-based content production and delivery.

2. Executive Summary

Object-based media (OBM) research and development is strong in the UK, both in academia and among broadcasters and content creators, with many of the concepts, tests, trials and experiments starting here. On an international stage we are often the crucible of development in this area. The UK has also pioneered many of the first commercial services. We have a strong cohort of start-ups and scale-ups in the area, as well as a rich talent pool of individuals who have graduated through the research departments of academia and the leading broadcasters and operators.

Our study however concludes that while most companies have financed extensive research in OBM, the transition to the availability of objects in mainstream content is currently very limited. In fact, many content distributors report that delivering basic objects such as subtitling, audio description (AD) and signing is a growing challenge rather than a shrinking one due the plethora of devices, platforms and delivery methods that they need to support today. Legacy platforms and methodologies cause considerable technological drag, a drag that newer OTT-only entrants suffer from less.

Web-based methods do however offer some real reason for optimism. The prevalence of browsers and browser-like functionality within operating systems, apps and hybrid platforms, means that the web's existing object-based text graphics and video tools used as part of web-page rendering can be used for the presentation of interactivity, graphics and other objects, as well as for user control. We see strong evidence that interactivity, personalization and the layering of objects are becoming commonplace on browsers and apps (e.g. YouTube, Amazon X-ray, Netflix and BBC Web Player). This does present a challenge and a growing disadvantage to devices or apps where web-like functionality is not well developed or is incompatible with that of a browser.

From a production point of view there are many very exciting developments such as Channel4 studio and BTs Social media team, who create a wide variety of "objects" that orbit their main programme output and augment the content on social media. We see narrative tools such as the BBC's Makerbox, as well as tools for building programmes with open OBM standards such as the IMF CPL Editor.

The UK is also home to many innovative and successful start-ups and scaleups working in object-based media, Endemol Shine and Stornaway have led in branching narrative objects, while Salsa Sound and BBC R&D are leading the way on object audio tools. V-Nova is working on layered video enhancement solutions. Targeted advertising is both a technical and commercial success for object-based media, with UK companies such as Finecast and Mirriad creating powerful partnerships with advertisers and broadcasters.

It is apparent however that a lack of consistency and standardisation in the way content is delivered from production companies to the operators is creating a situation where most of the production metadata which could be used for creating rich object based experiences is lost along the way. As a result, many broadcasters are forced to re-author lost metadata, such as scene information, credits location, and even subtitling, by re-analysing the completed linear

programme, either by hand or using artificial intelligence (AI) techniques. This results in lost efficiencies and additional overheads in extracting objects and metadata. Only next generation audio (NGA) and Object based audio (OBA) have the very mature and well deployed codecs and formats that allow for the delivery of object based media via broadcast. At least one NGA system is embedded into almost every new television (and many mobile phones, laptops and tablets too).

However, the palette of personalization features for broadcast-only content will always be limited as all objects need to be delivered to all receivers regardless of use or need. Currently many lesser-used objects are delivered to millions of viewers with no need for them, consuming valuable bandwidth. A channel with 5% viewership carrying an object used by 5% of their audience results in broadcast bandwidth for this object being wasted for 99.75% of receivers. For example, AD and subtitles are delivered via broadcast to all, whilst only a minority have them enabled regularly. As consumers move to online consumption, it becomes much more efficient to deliver only the required objects to each user and omit those not needed. This also opens out the possibility of not only filtering the objects delivered to each viewer depending on their needs and preferences, but also to pre-combine or cloud-render a personalised experience comprising only the objects needed for that user without delivering all objects to all receivers, as would be necessary for broadcast.

From a business perspective there are a few factors that are of note. Objects, except for targeted ads, have not yet proven themselves to have a strong enough business case to become “must do”. Accessibility, for example, still relies on regulation to ensure its growth. Object Based Media does however have the potential to tick many business boxes such as: production efficiency, audience engagement, enhanced user experience, reduced re-versioning costs, automated multi-format and multi-platform content creation, and increased ad revenue. As the benefits are multi-faceted, broadcasters and operators therefore need to make a complex internal business case for a shift towards objects. This is coupled to the unfavourable investment profile of object-based media where the initial effort and potential end-to-end change management needed to implement them is likely to take 3 to 5 years to implement and where the audience that can be addressed is still a small group. It is therefore not a simple ROI, and collaboration, support, as well as a few ‘carrots’ and ‘sticks’ are needed to encourage the industry beyond the current trials.

Overall, there are promising signs, with the most forward-looking broadcasters and operators developing innovative tools and processes that are producing better user experiences and production efficiencies. There is a very real opportunity, through initiatives such as The Prosperity Partnership and others, that the UK broadcast industry can come together and define new common methodologies and share tools for Object based media in a similar way to the DPP’s successful development of AS-11 in the last decade, which is used today as the standard delivery specification in production. Common agreement on such an object-based delivery specification would facilitate an ecosystem to develop which can utilise these objects in a multitude of ways that benefit consumers and operators.

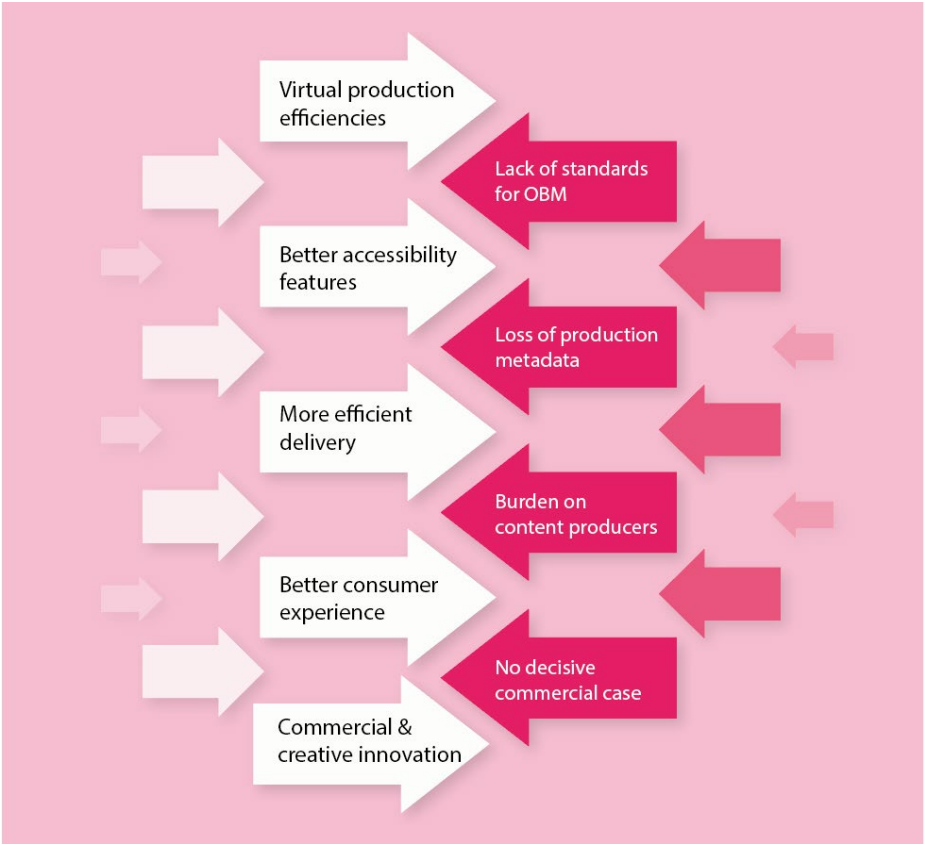


Fig 1. - Benefits and barriers to adoption of OBM

Key takeaways

1. Objects mean different things to different people. Two main categories of objects are emerging as the most prevalent; layered objects (which are presented together or as alternatives layers across content) and chunks or clips, (which are pieces of content that can be concatenated in different ways to present alternative experiences and narratives).
2. Some objects such as HDR, Audio Description, Subtitles, local news and targeted ads have resulted in the building of separate dedicated end-to-end bespoke workflows ONLY for the delivery of that object, whilst others are still waiting for a generic object carriage system to emerge.
3. Where to render the content is still a live and evolving issue. Today most content is rendered once in production, whilst in audio we are seeing some objects being kept separate and only finally rendered in the consumer's device. For cloud gaming and perhaps soon for media personalisation cloud rendering may increasingly be the solution. This however will only work where there is a bidirectional unicast broadband path available.
4. The most pressing use of personalisation made possible by objects is still extending accessibility. Where media providers are metadata-rich throughout their production to consumption workflows (at Netflix, Amazon Video and Disney+, for example) consumers are responding positively to the reliability of accessibility features for these services. Accessibility features are now becoming an aid to all, not only those with

specific needs. But only the giants are able to deliver consistency and reliability across all platforms and devices

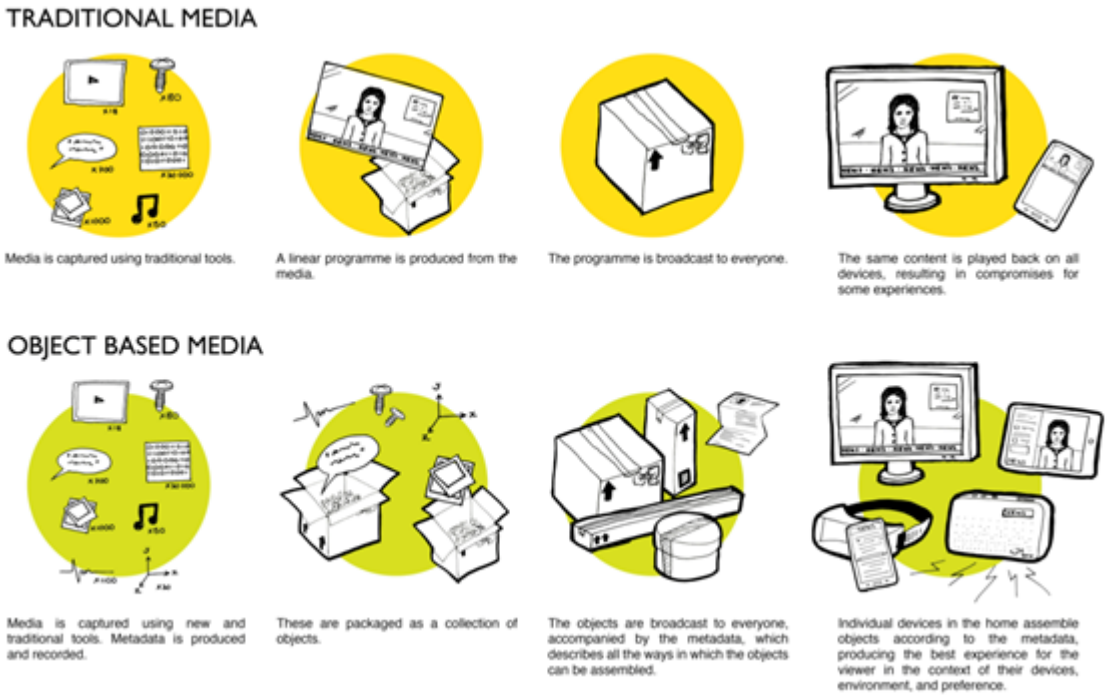
5. Objects and associated metadata are poorly defined in the production workflow, resulting in lack of consistency, loss of valuable metadata, and apathy from creators and operators
6. Audiences are already clipping, tweeting, personalising and distributing elements and clips from content on social media, usually in a homemade way on their mobile devices - operators and broadcasters are unsure how to react to this, either by providing richer metadata to make sharing and tagging easier for audiences, or by restricting object and metadata distribution to avoid inappropriate clipping and editing by consumers.
7. Broadcast and multicast only offer very limited possibilities for object based media (although NGA can be fully supported using a broadcast model). Hybrid or unicast offer far greater possibilities.
8. Point solutions designed for one type of object are hampering the development of a framework that can be used across all objects. This means that there are multiple highly successful single-purpose single-object workflows which solve each major use case individually (e.g. subtitles, AD, targeted advertising). This makes it harder to make the case for replacing these in favour of a workflow which caters for all object use cases.
9. Illustrating the benefits. The major companies interviewed indicated that in most cases they have very capable asset management systems in place that are capable of storing, grouping and tagging both chunks and layers of media (objects). However, in most cases only the minimum of metadata is populated, and media items stored. The reasons for this are that it is often seen as a tax on those who produce the content, and often not insisted upon by recipients. In cases where the value has been realised, such as sports broadcasting, the logging of metadata and tagging of key moments is both commonplace and highly developed (e.g. logging up to 15 items per minute relating to the action) allowing a plethora of personalization and searching and re-versioning use-cases to thrive.
10. Disrupt or be disrupted! Most major broadcasters also create and sell content to others. As newer entrants and newer platforms will quickly develop their object delivery capabilities (e.g. Amazon X-ray, Netflix HDR and immersive audio), suppliers of content to these platforms will be compelled to deliver media, metadata, and objects according to these platforms' requirements. This means that delivery formats such as IMF, and enriched object-based metadata may need to be in place. Retro-fitting and re-generating these is costly and time consuming compared to creating with objects in mind.
11. Where is the home of object standardisation? A common thread in the study is the need for the development of standards and tools that can be used across the industry, not merely home-grown separately by each broadcaster. There are several organizations working on these problems but none are dedicated to the topic. DVB and HbbTV have standardised TA, subtitling and AD, SMPTE have standardised the mastering formats such as IMF, and the DPP in the UK and EBU have developed other useful production standards in this area. Other parts come from the internet standardization bodies and industry fora. This causes a patchwork quilt of standards that pertain to OBM with no overall coordination.

3. Methodology

The term *object-based media*, as this report summarises, relates to a wide range of technologies where aspects of media production, delivery or consumption are organised as objects or elements. Objects are inherent in virtually all modern digital media technologies to some degree. Therefore, this report focuses on those aspects of OBM most relevant to major broadcasters, content providers and their audiences. The report consolidates findings from industry papers, blogs, software repositories and other available sources. In order to get a clear picture of both academic and industry points of view the report has taken a targeted approach to the subject, taking into account the context of recent reports and those commissioned in parallel by Ofcom, An archive of papers and reports from academia and industry was gathered to orientate discussion for a series of interviews.

It is also informed by a series of interviews conducted specifically for the purposes of informing the following report. Over a three-month consultation in February and April 2021, we interviewed 31 contributors. In this respect the report benefits from the insights and knowledge of relevant senior representatives from a number of major UK broadcasters, independent content producers, academic institutions and technology providers. We are grateful for the time and generosity of these contributors. Whilst we do not directly attribute findings to these individuals their names, alongside their organisations are listed in full in Annex C. Each interview was conducted using semi-structured questioning that allowed contributors to provide their own viewpoints on relevant aspects of object based media, whilst also allowing researchers the opportunity to pose emerging ideas and trends from the research with them. These findings are presented in the following report.

4. Introduction to object based media



BBC R&D illustration of traditional and object based media differences (Source: BBC R&D 2018)

What is object-based media?

Object-based media describes any approach to producing, distributing and/or consuming media content that uses separate digital media assets, known as 'objects'.

For example, a television broadcast of a football match might be produced as separate audio and video *objects* that would be combined during distribution to be consumed as a singular media experience by the consumer audience. By keeping these objects separate, media content can change according to their requirements: a viewer watching a football match might choose to hear commentary from a pundit who is a supporter of their team, instead of the pundit from the standard broadcast or a pundit from the other side. By producing three audio objects, one for each team and one neutral commentator alongside one video object, choice can be offered to the viewer. The viewer can interact by making explicit choices through user controls (e.g. selecting a particular commentary from the list of options on screen), or implicitly through data provided by the user or their device (BBC, 2019). As well as offering new opportunities to adapt the narrative experience, object-based media also offers opportunities to optimise the presentation of content, and offer content experiences driven explicitly by the user (BBC, 2020).

To further illustrate the concept of objects let us look at a typical broadcast programme. Even the simplest programme typically contains several media objects, such as audio, video, subtitles, an electronic programme guide (EPG) listing, and sometimes signing and audio description (AD). Most programme packages will also include a trailer, bumper, promo photos, longer EPG listing, and other elements. Even within the delivered programme itself there are multiple elements that can be delivered as objects such as the opening credits, closing credits, the theme tune, as well as objects within the audio and video itself such as left channel audio,

right channel audio, (and other channels if it is produced in surround sound). The video itself is also delivered in multiple versions: typically, HD, SD and sometimes UHD. The video may also be chunked up into sections or chapters, including opening and closing credits and sections between advertising breaks.

As media production, post-production and media delivery evolve, the number of objects associated with the content increases. Social media trailers, promotional web content, images, sound clips and a host of other assets are now associated with most media programmes. All of these elements have traditionally been managed in a fairly *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/storage media. Well organised object based media uses metadata to describe the properties, relationships between, and uses of objects. This information allows complex packages of media to be adapted and rendered dynamically and at scale.

Subtitles, Audio Description, Signing, and trailers were often not stored with the content and needed to be re-authored when the content was to be aired again. Other industries have however taken a different approach to media and the concept of objects is an integral part. The Games industry has been “object based” from the outset with games engines used to play-out or render these elements depending on the gameplay. Thousands of snippets of data, pictures, audio and video are stored as part of the game and combined on-the-fly as the game unfolds. These game elements or objects are always the same, but the way they are put together is different each time the game is played.

Accessibility and personalisation: the history of object-based media

Although the term ‘object-based media’ is associated with digital content, it can be understood as a part of a steady evolution in conventional media spanning many decades, in which elements of a broadcast are produced separately and combined together during playback depending on the requirements or technological affordances of its audience. Similarly, subtitles for the deaf and hard of hearing were originally burnt into visual assets. Teletext subtitles which first appeared on BBC Television in the UK in 1979 could be regarded as the first true “object” because although it was carried with content, it was possible to select or deselect whether the subtitles were shown on the screen. Live subtitles were first broadcast in 1984 and by 2008 the BBC had achieved 100% subtitling for all of its main channels. Since 2012 the BBC iPlayer has provided 100% closed caption “object-based” subtitling on all capable platforms for its on-demand and downloadable content. By now subtitling on some platforms is offered in multiple languages, character sets and colours and can be repositioned on the screen through the use of metadata.

Whilst initially closed captioning was intended for the hard-of-hearing, today it is popular in noisy environments, for media consumption on the move for example, making core accessibility features increasingly essential for a much wider range of consumers.

Another optional accessibility feature is audio description and both features are examples of objects, layers of the broadcast that can be turned on and off by the viewer (see ‘Common types of object’). Object-based media extends the possibilities for accessible content providing

much higher levels of personalisation for each viewer. The most popular of these is the inclusion of signing, where Ofcom is working with broadcasters to increase its availability. Objects allow for the inclusion of picture-in-picture style signing video objects for flexibility around the placement and size of the media object. Audio elements can be orchestrated to provide clearer audio, (see Salsa Sound service description p28,47). These scenarios are made possible because the content is created as a number of components which are assembled at the receiver and consumed according to the user's needs.

Common types of object: layers, chunks and transmedia

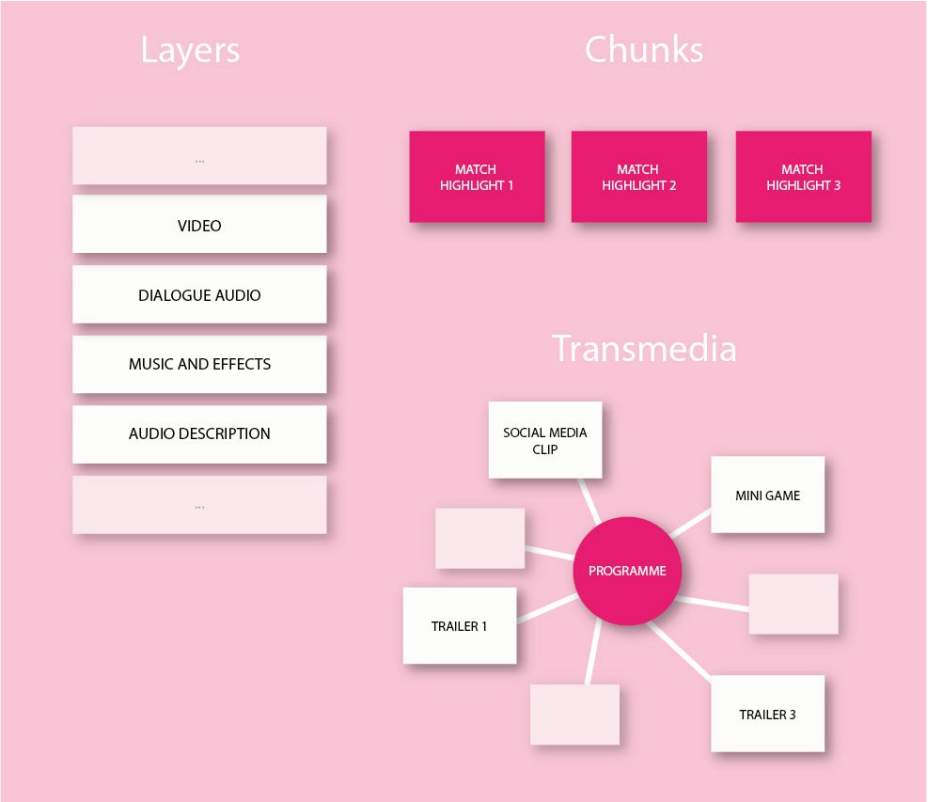


Fig. 2 - Layers, chunks and transmedia objects

Layers

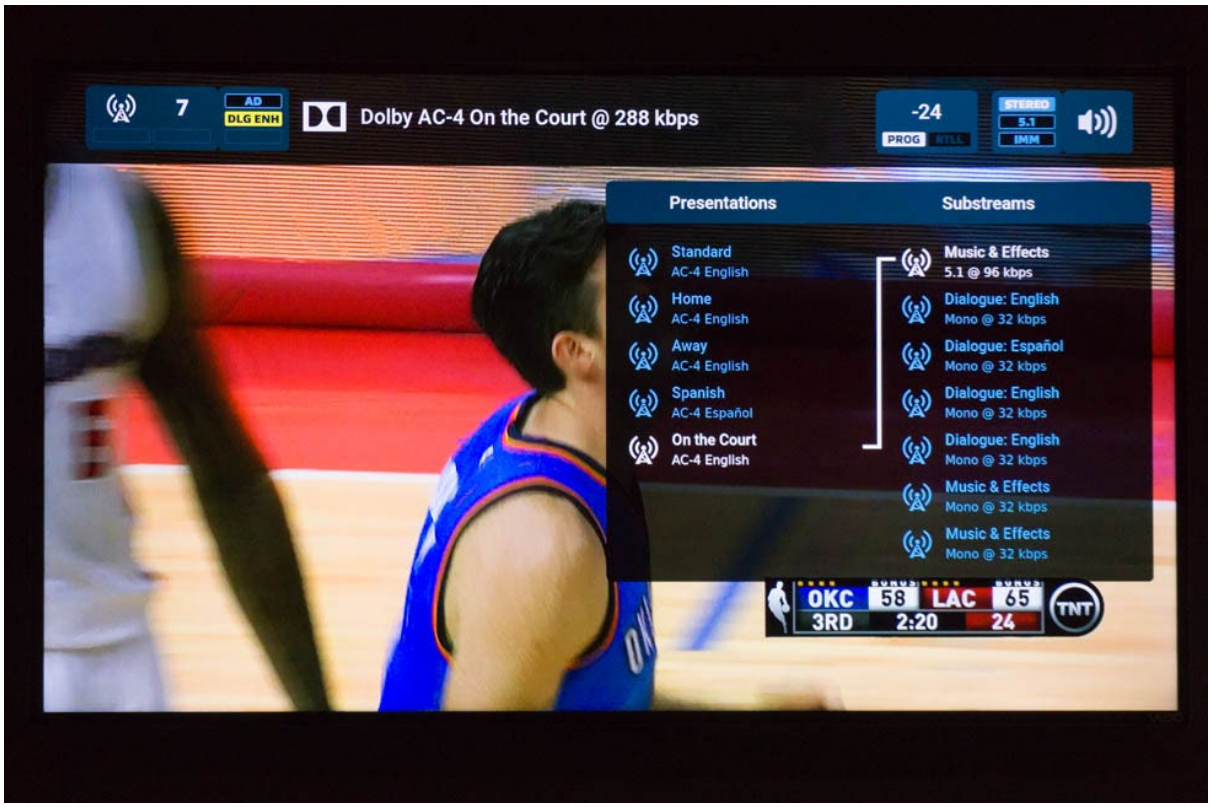
Layers are objects either designed to be consumed together with or instead of other objects at a moment in the content timeline.

Common examples include:

- Subtitles and audio description
- Director's commentary
- Alternative graphics, commentaries
- Alternative languages
- Separated sound or video effects (e.g. to change dialogue loudness for better clarity)

Objects of this type are combined in groups to create alternative or augmented experiences. These objects must be co-timed and synchronised to be meaningful.

Example:
An object-based audio experience in which 7 audio tracks (layers) are delivered to the consumer and combined in different ways to create 5 discrete audio presentations in English, Spanish, Home and Away commentators and a commentary-free version



Showing the different audio objects translating into consumer options (Source: Henninger, 2016)

Chunks

The other type of object which is developing in both application and popularity is the application of object media clips or *chunks*. These objects are a sequential segment of a programme. Examples of clips include:

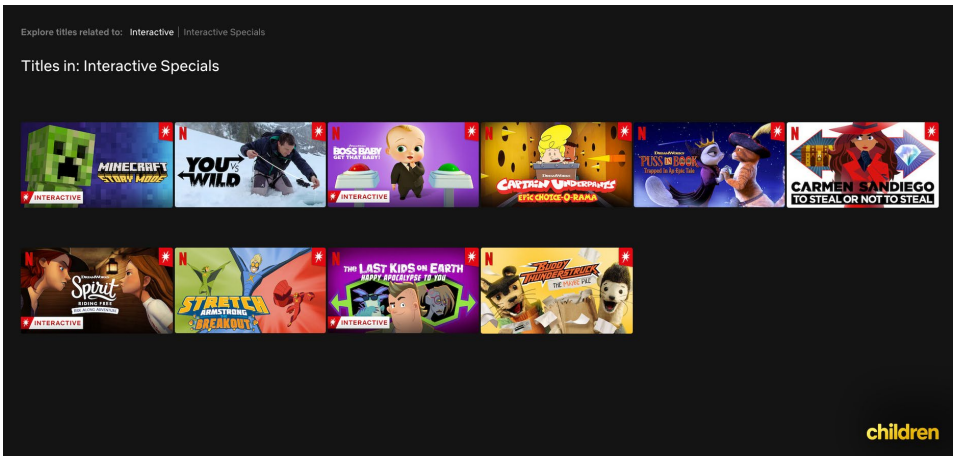
- News items
- Boxing rounds
- Scenes
- Opening or closing credits

Manipulating these chunks gives consumers and creative producers more flexibility in terms of what audiences consume, allowing the production of:

- Different programme lengths for different production or consumer requirements (e.g. from episode highlights to full episode).
- Branching interactive narrative that allows audiences to make choices that affect narrative outcomes.
- Media tailored to the individual by creating new content packages from a long list of content (e.g. personalised news package for people who opt out of celebrity news).
- Targeting advertising at specific consumers with specific ad clips.

Example: Netflix Interactive: You Vs. Wild (featuring Bear Grylls)

Following the Black Mirror Bandersnatch interactive episode, Netflix continues to trial its interactive offering with interactive children's programming with its Interactive Specials series:



You vs. Wild is part of a wider series of Interactive Specials (Source: Netflix, 2021)

One of which is the specially commissioned You vs. Wild in which young viewers can help Bear Grylls make the right survival decisions on a rescue mission in the jungle.

Transmedia objects

Transmedia objects orbit 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 objects include bumpers, trailers, making-of documentaries, outtakes Wikipedia/IMDB entities, cast and crew interviews, location information, reviews, social media pages, blog posts etc. They can even extend to a related radio drama version, toys, merchandise, magazines and spin-offs. Sky's Immersive VR app with augmented content is a good example, as is Amazon's x-ray feature.

All of these objects exist today but are very rarely arranged, ordered, tagged and searched as a true object database.

It is however very feasible to create such a rich object-based database and metadata which allows this orbiting cloud of objects to be linked to the content in ways which benefit both consumer and content provider.

Example: Netflix HyperPersonalised UI

Today Netflix uses AI to fully personalize the still photos that are used to promote each programme in the user interface and VOD gallery. This method uses a list of transmedia objects (in this case a database of promo photos associated with each programme). The UI then displays the photo that the AI believes is closest to your programme tastes, increasing the likelihood that the user selects that programme.

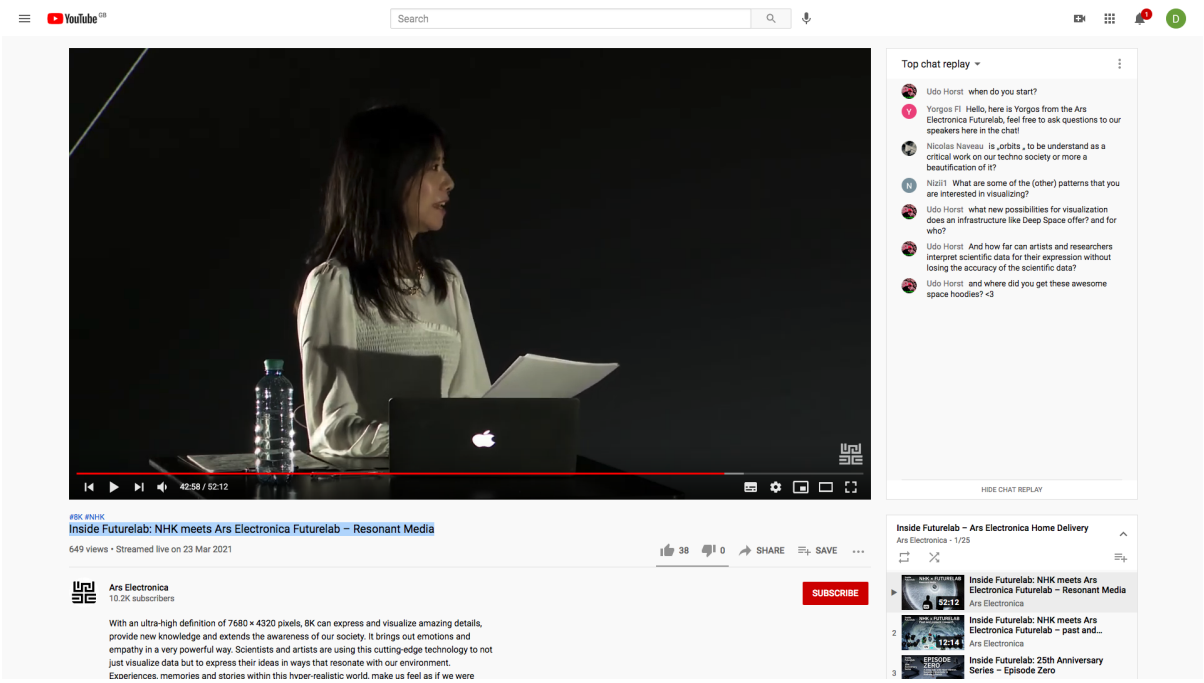
Example: Amazon X-Ray adds additional detail through OBM

Amazon X-Ray for Prime Video provides an additional text and image layer for audiences, tailored to provide information about on-screen subject matter. This can include background music, character and actor information and trivia about the episode or series.

Amazon X-Ray service provides details of onscreen characters and actors. (Source: Amazon.com)

Example: YouTube Live Chat Replay

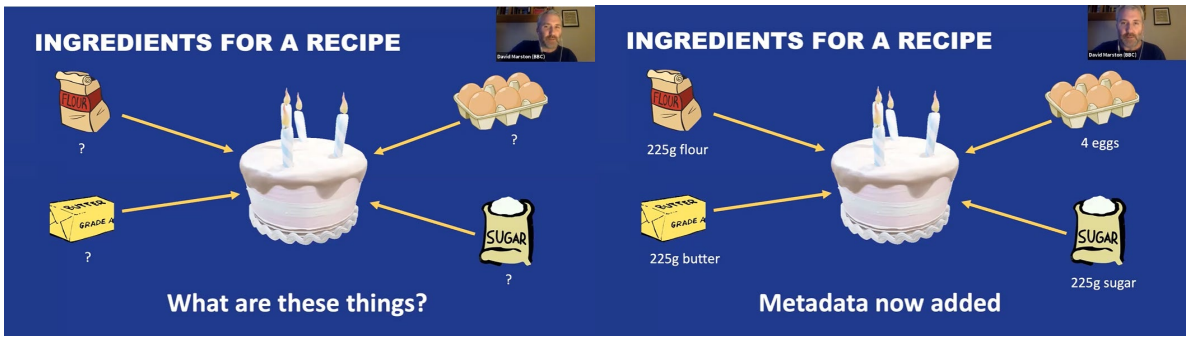
An example of this can be found on YouTube's live events streaming service. During the live event, audience members can comment on what is happening and these chat items are stored as objects and played back in time with video media when it is replayed, providing greater context and an enhanced sense of 'liveness'.



Live Chat replay next to Ars Electronica event video on YouTube (Ars Electronica/YouTube, 2021)

Metadata and objects

Objects alone hold no purpose unless they are accompanied with metadata to explain if, what, where and how they can be used.



Dave Marsden (BBC) explains the importance of metadata in an EBU beginners guide to ADM (Source: Marston, 2020)

It is perfectly possible to create a single programme with little or no metadata. A title will however help that programme to be found, a synopsis will help it to be catalogued, and a cast list and genre will help place it within the canon of other content. In OBM, classes of metadata are also used for building the media itself. For example, narrative metadata might explain the relationship of this scene to others e.g. must be before x but cannot be after y, and introduces character z. Audio metadata will accompany each and every sound object in a scene and can include the position of the sound, its volume, loudness and intensity, whether it is selectable or not, and rules of use (e.g. you can select the Welsh audio description but only to accompany the Welsh language narrator). Video metadata might include camera angle data, tags for the person in a closeup, and even AI or human generated *descriptions* of what can be seen. With all these pieces of metadata in place new possibilities would be opened to re-version and navigate content, to create highlights, automatically create trailers and bumpers, even to do this on-the-fly, specifically tailored to an individual's needs. In this context, metadata becomes all-important because of the need to create relevant tags to better define user preferences. Whilst it is common for humans to add metadata to media, machine learning is increasingly playing a key role, processing large amounts of data in a meaningful granular level. However, this automation is often put in place to recover metadata that was present but was lost or stripped from the media object within the workflow.

Metadata loss during production and post-production

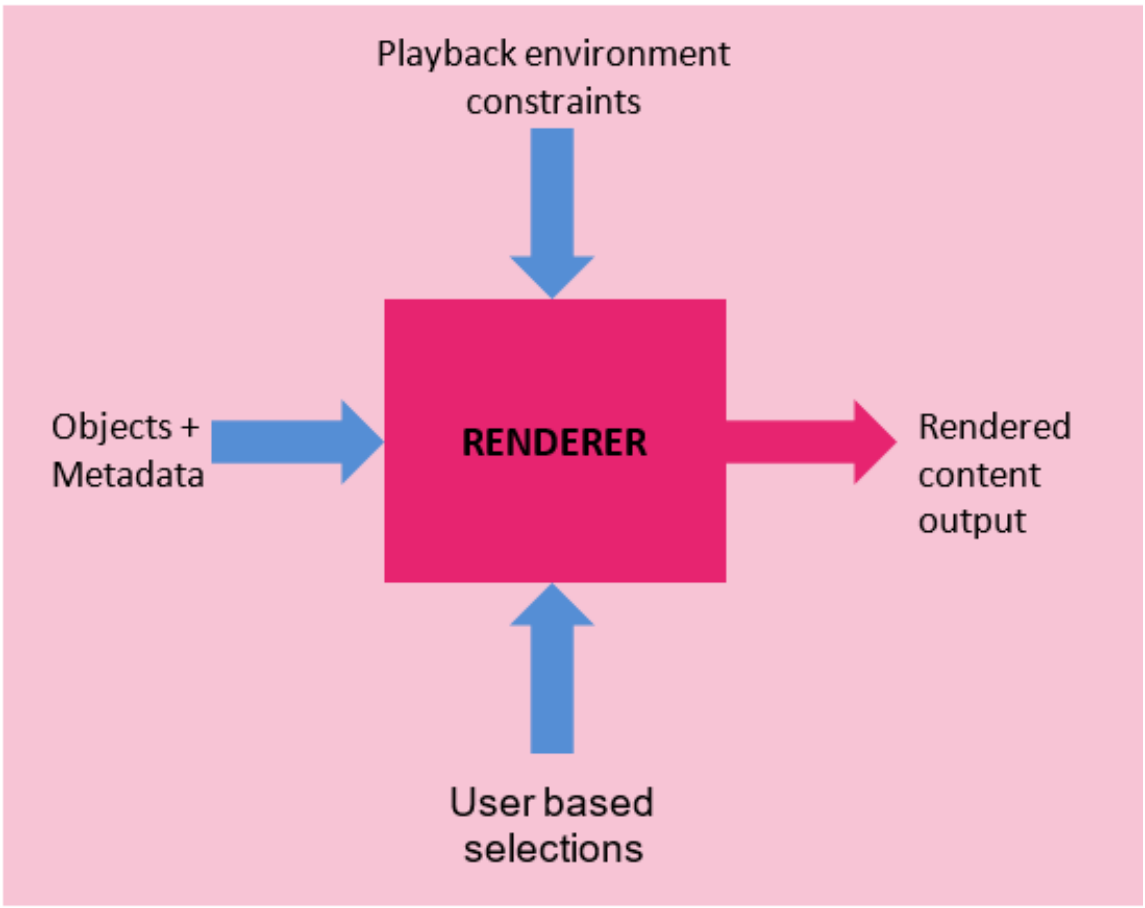
Whilst there is currently an abundance of metadata associated with broadcast media objects at the production stage that could be used to power more dynamic distribution and consumption models, for many mainstream UK broadcasters there is very little standardisation around capturing this. The existing AMWA AS-11 specifications do not require the capture of this scene-by-scene data. Therefore, research and development teams seeking to develop projects in this area are required to retrieve metadata at a later stage with more costly and time-consuming artificial intelligence approaches. Whilst there is a growing number of effective automation tools in this area (see 'AI and metadata automation'), these are generally for

specialised applications such as live events. Other broadcast formats such as drama are harder to standardise from an AI perspective.

A standardised production requirement related to object-based media tagging (effectively gluing the production data to media) would help to ensure that less data was unnecessarily lost from the production process, and standardisation would offer interoperability for production houses across both public, commercial broadcasters and OTT service providers. The Interoperable Mastering Format (IMF) offers a good start here (see section - Open standards for OBM), but a shift in attitude towards the preservation/creation of metadata is needed which will only come when the value of it is clearly articulated.

Rendering: turning objects into media

The final concept of object-based media to understand is rendering. This term is not new and has been used in the visual effects world and in the games industry for many years. The purpose of the renderer is to assemble all the elements into the flattened final video and audio ready to be consumed. In object-based media there are more elements that need to be considered by the renderer, but in essence it combines some or all the objects available using the accompanying metadata combined with constraints imposed by the playback environment and the constraints or selections made by the user or made on behalf of the user based on preferences.



Multiple inputs inform what is outputted by the renderer

Playback environment examples include:

- size or type of device or screen
- number and position of speakers
- use of headphones
- ability to display UI or graphics

User-based constraints include:

- Closed captioning choice
- Dialogue enhancement
- Signing
- Language preference
- Age restrictions (in relation to violence/ nudity/ profanity)

The concepts are similar for rendering audio, video or graphics. For example, If the renderer knows that there are accessibility needs for size of text or addition of signing, it will take these into account, using the objects it receives to create the final displayed video accordingly. If the content is broadcast/multicast only (one to many), the renderer is the only way that the consumer has to make changes to the playback.

Renderer positioning in the content workflow

There are three principal places where the rendering can take place in a linear narrative, and a fourth which relates to Branching Narratives:

1. **Production Rendering** by the content creator during the production process. This is the traditional method where one or more versions of the programme are made in production and delivered as-is to all consumers. This is a one-size-fits-all approach, although in practice programmes are rendered (versioned) many times in production to cope with different quality levels, platforms and markets. This is currently the most common form of rendering for the majority of broadcasters.

2. **Cloud/Edge Rendering** in the cloud or at the Network edge. This method is akin to cloud gaming where all available objects are delivered and stored in the cloud. When a consumer wants to watch the programme their preferences inputs and playback environment are uploaded to the cloud and the content is rendered according to these instructions specifically for them.

3. **Device Rendering** in the consumer's device. The objects and metadata are delivered to the consumer and the renderer in the consumer's device builds the content playback locally based on user preferences and playback environment.

4. **Rendering Branching Narratives**. Rendering in a branching narrative context is rather easier, as branch points do not require objects to be combined during playback and only the choice of which objects come next is affected. Cloud-based rendering (which in essence is just a modified content playlist) is the only feasible way to achieve this, local rendering would need all the content to be downloaded in advance and be

available at the time of playback (akin to DVD scene selection, for example the 2000 Ultimate Edition DVD of *Terminator 2: Judgement Day*, the 2007 DVD rerelease of *Blade Runner* in which three different cuts of the films are playable from the same disc.). YouTube, Netflix and many other app and browser-based platforms now support these branching narrative features and tools such as StoryFormer and UK start-up Stornaway are making it easier to create, export and play-back the content.

Production rendering

In traditional production content rendering is firmly in the hands of the production and post production teams. All the elements or objects that have been created, clips of video, audio elements, graphics, voiceovers and visual effects, are gathered and combined to create the final programme which, once made, cannot be altered, unless and until another fully rendered version is generated.

User preference elements and the playback environment elements can only be accommodated by using a one-size-fits-all approach whereby the programme makers create the programme in a way which is acceptable to all viewers, but specifically tuned for no one. This often leads to compromises in all aspects including the editorial and narrative of the content itself and its technical formats.

It also leads to multiple versions of the content having to be created, stored and distributed to cater for cases where one-size cannot be made to fit all. Examples of these cases include a signing version, an HD and an SD version, a version for before and for after the Watershed, or an omnibus edition.

There are many examples in which production rendered versions of programmes number many dozens, with some platforms talking of having hundreds of versions of the same content. Creating, storing and keeping track of these can be costly and time consuming. The corollary of multiple versions is a generic version which is a compromise for all who consume it, as it has to take into account the lowest common denominator for all aspects. The side effects of these issues are that some versions that are only relevant to small audiences, e.g. the signed version, are often not created at all, or have limited availability.

Most content is still made in this way today, with the exception of subtitled versions and the audio for some content which is delivered in multi-channel format and downmixed locally for stereo devices.

Cloud-based ‘Edge’ Rendering

This can be considered as a half-way house or even as an ideal scenario from a rendering perspective where the specific version delivered to the consumer is tailored to their needs in the cloud or at the network edge. This method requires a bidirectional path to the consumer as the preferences need to be uploaded to the cloud in order for the renderer to take these into account when creating the final version.

Examples of the preference data that could be used by the cloud/edge renderer would be Subtitles, AD, signing preferences, age, location, type of consumption device, size of screen, number and type of audio speakers or headphones, ambient light, ambient noise, emotional state, narrative preferences and so on.

The sum of this data can in some cases be very revealing of many aspects of the consumer's life and careful treatment of this data would be needed. Just like the tailoring of bespoke clothing, personalization of media, by its very nature, needs intimate knowledge of the customer to provide the best result. And trust is needed to allow this to be shared.

Cloud or edge rendering is simply not possible in a broadcast/multicast scenario. Even where a bidirectional broadband connection is available, it can be expensive and complex to implement, maintain and run, as the content provider would need to make all objects available to the cloud/edge, including storing them there, and have the network capabilities to render them for each consumer via cloud servers.

It can also be expensive from a last-mile bandwidth perspective as each consumer would get the best version their device connection can process, which is likely to be a much higher bandwidth/better quality version than the lowest common denominator version they get today.

Consideration is also needed for network latency, as any change in preference or input from the user will take time to be uploaded to the cloud, so that the cloud render can make the changes to the content. There is then further delay before this altered content is sent to the consumer, and they see the effect of the changes made. These delays are not material for user choices such as adding subtitling or AD.

On the positive side cloud/edge rendering does provide a very compelling way for everyone to get the best personal content experience tailored for them without needing complex and expensive rendering in the device or having to deliver all objects to all consumers.

Device Rendering

For device rendering to work all the relevant objects and metadata have to be delivered to the device, the device will then combine these according to the information it has about the user's preferences and playback environment, as well as the information it has about its own device playback capabilities.

This technique works well for both broadcast/multicast and broadband/unicast delivery, as no return path is needed.

In a broadcast scenario, all objects would be delivered to all consumers and would be combined to taste on the device. Next Generation Audio can work in this way and the objects are usually all carried within a single stream.

If a bidirectional connection is available (Hybrid or Broadband) then additional objects can be requested and delivered on-demand, again to be combined locally.

From a cost and complexity perspective this method places a higher burden on the device and, often, bandwidth because in a broadcast/multicast scenario, elements and objects are being delivered to consumers which they will not use in their rendered version of the content.

Hybrid delivery also has its challenges as multiple delivery paths for different elements means a complex and highly robust method of re-syncing the objects at the receiver is needed.

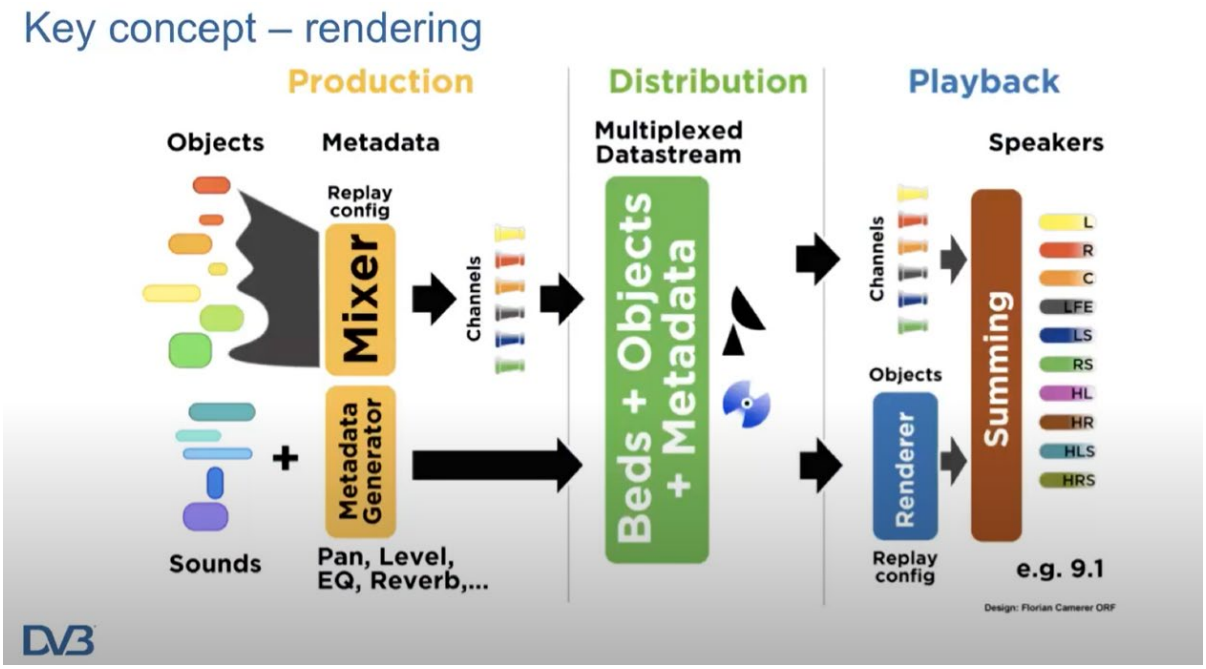
In an unicast scenario, if the local renderer only requests the objects it needs at a given time, the bandwidth can be reduced, but at the cost of increased latency, as when the user makes a change (e.g. switching on graphics or AD), the request to deliver additional objects needs to be made to the network and these objects need to be delivered and synced before the renderer can add them. This can result in slower responses to changes made by the user.

Browser-based Rendering

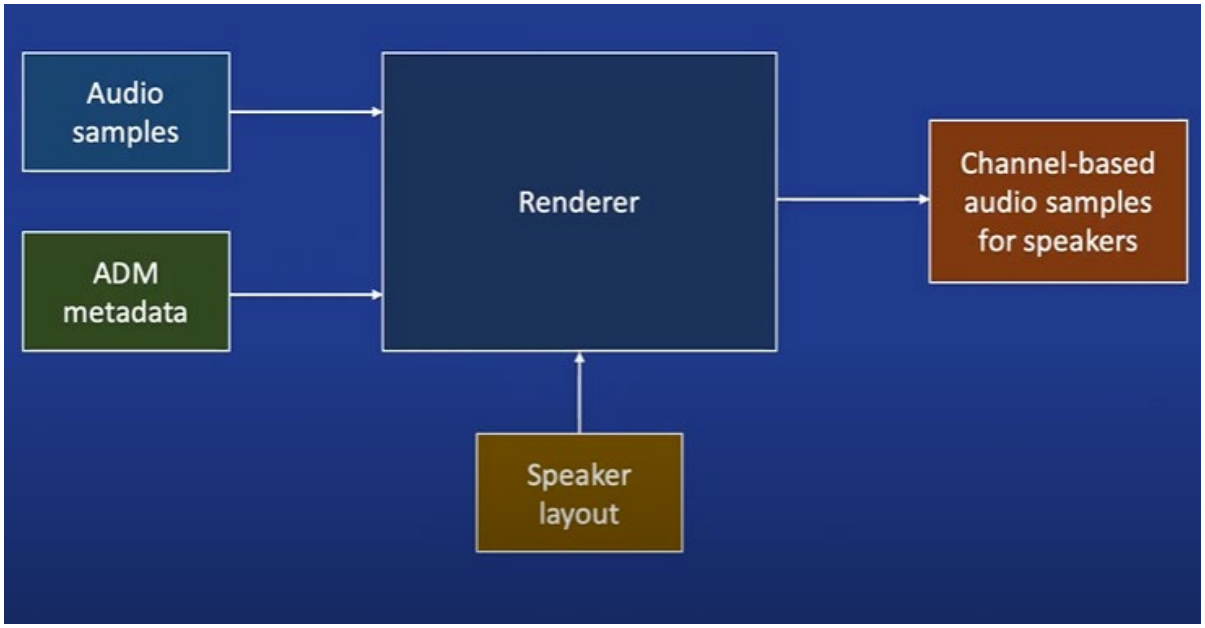
As content consumption is increasingly taking place on devices which include browsers or where the browser itself is the playback mechanism, the possibilities for rendering open out considerably. A browser by its very nature is a rendering engine and browser content is by default object based. Consider visiting any web-page today. Your content experience is already fully tailored and rendered specifically for you. The ads you see are personalized, the way the content looks adapts automatically depending on the device you use, and size of the window, it takes into account your accessibility needs and your user preferences. This is therefore an ideal rendering engine for content and is already available almost universally on all new media consumption devices. No wonder, therefore, that interactivity and object-based content is gravitating this way. For example the BBC web player had additional interactivity features compared to the iPlayer and many of the experiments we have seen (CAKE, select a quest https://www.selectaquest.co.uk/) use the browser or web-app to manage the rendering. It is therefore no surprise that HbbTV, BBC and major broadcasters look to lobby W3C and IETF for standardization of media manipulation and rendering rather than turning to traditional broadcast standards.

NGA Audio rendering

In most cases, Next Generation Audio codecs and technologies rely on local device rendering. This reliance allows their use in broadcast scenarios. As receiver mixed Audio Description and downmixing from multichannel to stereo were already commonplace even before NGA, it was a natural extension to add further object-based capabilities to these devices. The technologies themselves are also geared towards device based integration and licensing and therefore the device is the natural place to include the technology.



Notes on rendering from DVB webinar 'Next Generation Audio in DVB' (DVB Project, 2019)



ADM/EAR renderer for Audio explained in EBU article 'NGA and the EAR production suite' (P.Sunna EBU, 2019)

The capabilities of these local renderers are however limited, and in many cases, content is delivered using a method in which generic audio elements are grouped together and delivered pre-mixed and only the selectable or manipulatable elements are delivered as objects. This technique is useful and effective for delivering, for example, the sports stadium sound as a pre-mixed 5.1 format which can be down-mixed to stereo or binaural in the consumer device. Device based rendering with a mix of individual objects and pre-mixed objects therefore provides a realistic path to the benefits of NGA for the broadcaster, using capabilities already commonplace in devices. This can be seen in the NGA experiments conducted by major operators in the UK and elsewhere.

For internet delivered content, cloud-based rendering of audio is however possible and often preferable. Selecting a language track on an OTT platform simply selects a different full track to stream rather than replacing objects. In the Games world, audio has been treated as multiple objects for a long time, with the game's engine combining/rendering the objects from a scene on-the-fly as the user plays the game, personalizing it to their gaming experience. No NGA systems are needed for this, as the audio is not compressed or transmitted. For cloud gaming, the same principle applies but the rendering is done in the cloud and only the final (pre-rendered) audio for each specific user is compressed and transmitted.

Rendering Branching Narratives

Rendering in a branching narrative context is rather easier as branch points do not require objects to be combined during playback and only the choice of which objects come next is affected. Cloud-based rendering (which in essence is just a modified content playlist) is the only feasible way to achieve this. Local rendering would need all the content to be downloaded in advance and be available at the time of playback (akin to DVD scene selection, for example the 2000 Ultimate Edition DVD of *Terminator 2: Judgement Day*, the 2007 DVD rerelease of *Blade Runner* and the Platinum and Diamond Edition DVDs of *Beauty and the Beast* in which three different cuts of the films are playable from the same disc.).

Example: A Heist with Markiplier is a branching narrative YouTube Original using the play next feature of the platform.

The viewer plays a part in an comedic interactive narrative (Markiplier, 2019)

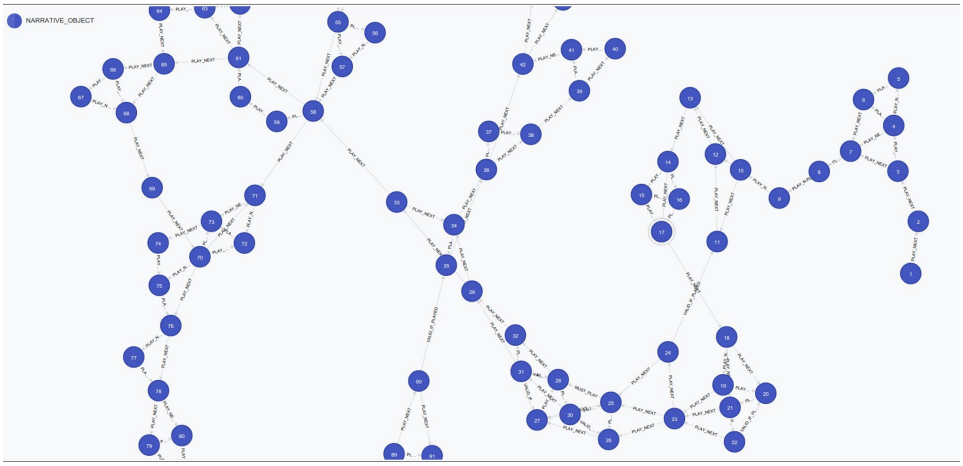
Content selection

If the content is delivered as a unicast (*broadband*) stream or the full object database is available locally to the consumer other dimensions open out such as nonlinear playback of the content. This includes playing news items in a different order or creating a very short or a longer version of the programme according to the user's request.

Interactive elements can also be included as illustrated by Netflix and others (see Common types of object), creating different routes and branches in the narrative

One of the best early examples is the BBC R&D's responsive radio demo (BBC R&D, 2015) in which the length of the programme could be seamlessly varied to suit the listener. In this example, the consumer is never aware of the branching narrative, never needs to make a choice beyond the original "length" selection and is completely unaware of how other versions differ. This is a fully automated branching narrative driven, in this case, by the selected length.

Automotive manufacturers are apparently very interested in this kind of media adaptation as programmes can be tailored to the length of a car journey to keep passengers entertained or adapted to keep the driver entertained while the electric vehicle re-charges, the programme adapts to fit the charging time so that the driver is entertained while waiting for a battery top-up.



This diagram shows the objects (media chunks) within the programme as blue spots and the various paths through the content as the connecting lines. (Source: BBC, 2015)

5. Technology trends in OBM

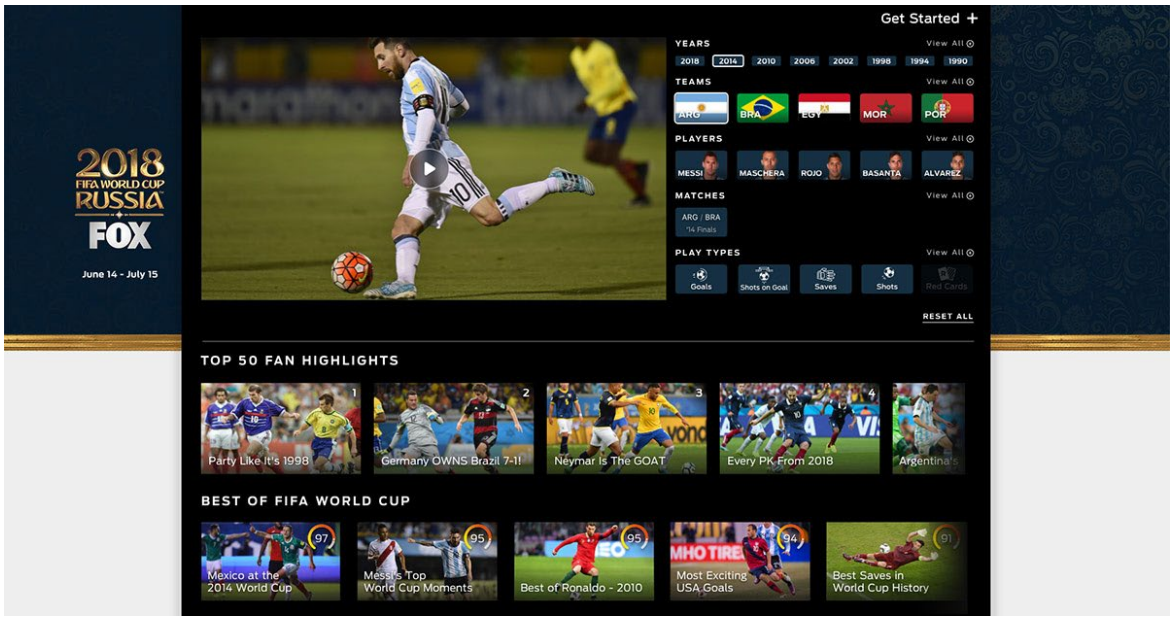
IP-only and hybrid broadcast

Higher broadband penetration is increasing viewership on mobile devices and laptops alongside IP-only and hybrid TV where content providers such as BT are offering enhanced sporting spectator experiences such as BT Sport Player. Netflix are investing heavily in user-centred narrative experiences such as *Bandersnatch* and children's content such as *You Vs. Wild* which take advantage of a return path from the viewer's device to provide interactive narratives. Hybrid televisions are also providing some OBM functionality via NGA and HbbTV features (PwC, 2019; Ofcom, 2021). These technological trends are being driven strongly by changing consumer habits during lockdown as consumers seek more quality content on an on-demand basis.

Machine learning (ML) and artificial intelligence (AI)

Developments in ML/AI are increasing efficiency across the production, distribution and consumption chain. In particular, new AI services are able to bring down the cost of OBM production through the automation of media object creation in recorded footage, alongside a growing set of automated editorial functions (e.g. IBM, 2018). The growth of OBM is likely to be reliant on the increasing ability of these tools to move into a wide-spread business-as-usual setting.

Example: IBM Watson Media and Fox Sports



IBM Watson facilitated interactive media coverage for Fox Sports (Source: IBM Watson, 2018)

For the 2018 World Cup, Fox Sports worked with IBM Watson Media to augment their broadcasting processes through AI analysis of media objects, people, words, concepts, emotional overtones and character personalities to deliver more relevant content to consumers and add efficiency to video production and editing. The tool had previously been used by IBM in collaboration with broadcasters at The Masters, Wimbledon and the US Open.

Their cloud-based video AI service branded Cognitive Highlights provides a video metadata enrichment service that allows video editors to more quickly package and distribute highlight reels by automatically identifying and ranking exciting match moments.

Microsoft, Google, AWS and Huawei also offer similar cloud-based services. AI and ML techniques now form a part of many content workflows.

Example: Salsa Sound

UK company Salsa Sound is also working with object-based media to increase editorial speed, using audio objects to automate aspects of sports editorial. Its neural-network based algorithm uses existing pitch-side microphone feeds to track sound events on the pitch, with no video or other tracking required. It uses this information to enhance on-pitch sounds for better audio using AI. Salsa can locate the position of on-pitch sounds to within around 0.5m, outputting the data in real time. Important events such as free kicks, penalties and corners are tagged with time and location metadata and generate a live datastream suitable for use in automated camera control, highlight package production and the automatic triggering of on-screen graphics.

Our research finds that most UK broadcasters are using a combination of in-house or external artificial intelligence services and teams to increase efficiencies around object metadata retrieval and editorial automation.

Games Engines and virtual production

Game engines are playing an active role in both production and increasingly in the playback of content. The biggest video games such as Fortnite and FIFA Football rely heavily on media objects, fragments of recorded audio, imagery and sometimes video which are conditionally combined during gameplay. The line between entertainment content and games is blurring at an increasing rate, with TV and games sharing production tools and esports - watching expert video game players play and compete - now a significant segment of online viewership (MBA@Syracuse, 2021).

In fact, game engines are becoming a mainstay in virtual production where locations and special effects can be simulated on high-definition LED screens or greenscreen during filming. US company Epic Games, originally a video computer games maker, provides Unreal Engine, a game engine that is now becoming a leader in end-to-end production tools that provides an interface and real time renderer for object-based media for film and TV production. Realtime 3D assets in production settings backdrops are currently used in a number of virtual studio environments including Match of the Day where Epic Games' Unreal Engine is used to render a virtual pitch side studio which adjusts realistically to camera movement and shot changes.

BBC Sport personalities in virtual studio based in Salford Quays (Source: Dock10, 2020)

Games engines powered by media objects are also being utilised to create animated children's programming as they offer a very fast and cost-effective way of creating CGI animation and fun or educational graphics.

The demand for games engine specialists is only increasing, sparking a growing demand for hybrid specialists in the UK with skills in both games engine usage and TV and film production and a skills and experience gap in the sector (Bennett & Murphy, 2020).

Epic considers reusability of objects from the virtual production workflow to be a key point of difference when using Unreal Engine.

> "You can use the same assets for interactive and location-based entertainment, marketing materials, and more."
>
> (Epic Games, 2021)

The use of game engines in the way that Epic imagines, suggests a possible growth in transmedia object development where connections between objects are not organised simply around audio and video playback but around reusable assets in a storyworld. In this way game engines provide new ways of adding value for the consumer, new channels for monetisation and increasing efficiency for producers and content providers such as broadcasters and games companies.

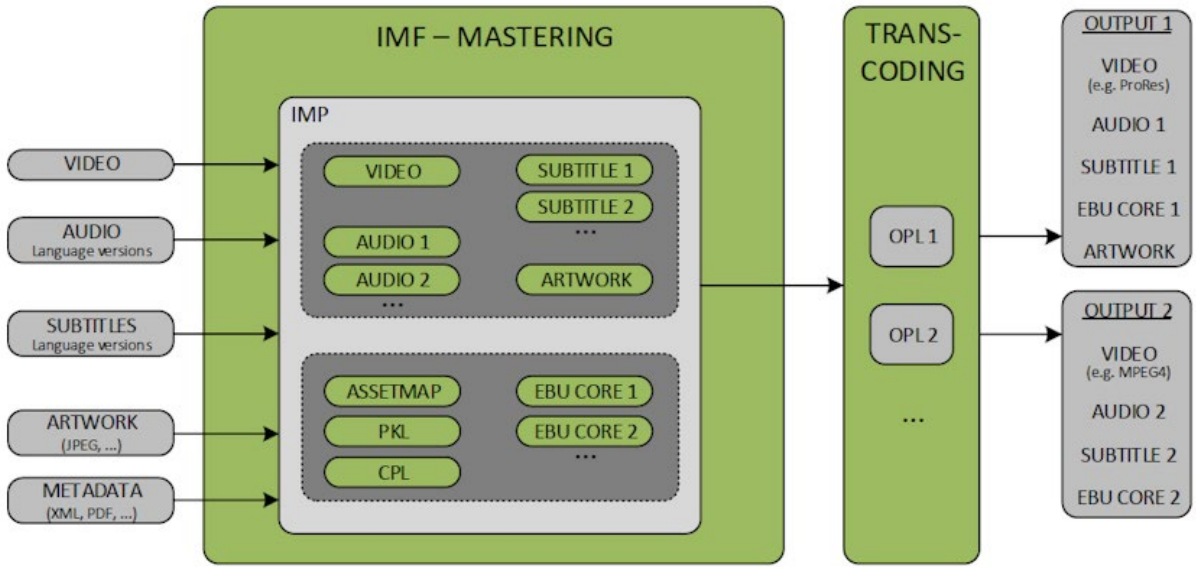
Open standards for OBM

In order for OBM to leverage greater efficiencies, interoperable standards are essential. Two major open standards for digital broadcasting are currently in use that provide different levels of object-based media flexibility and practicality.

IMF: the working OBM standard used by Netflix and many others

The Interoperable Mastering Format (IMF) is an object-based media standard that allows the creation of IMF packages which comprise media files alongside corresponding assembly instructions. Video-on-demand leader, Netflix, adopted IMF in 2014 and brought its OBM standards for OTT post production workflows into the mainstream. Unlike other media formats, video media is not delivered in one file comprising audio and video together. A typical IMF package media package contains:

- Video essence (J2K up to UHD)
- Audio essence (24bit uncompressed, any number of channels)
- Data essence (subtitles and captioning using IMSC Timed Text)
- Dynamic metadata (metadata that changes over time)
- Composition playlist (CPL) – human readable XML
- Packaging data XML (asset map, packing list and volume index)

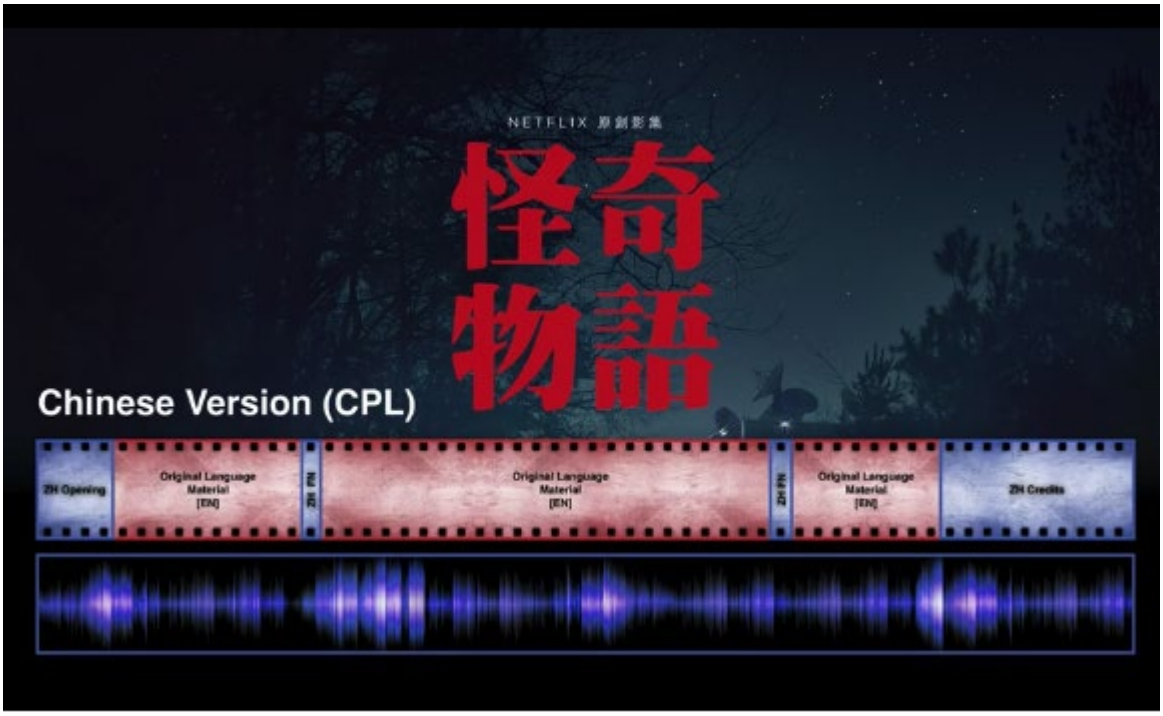


It is the CPL file that provides the instructions for compiling and playing the content, ensuring that only the content that is required is rendered. This is particularly useful for versioning of content for different territories, where the majority of video content stays the same except for credits which may be switched to the home language of the viewer, and a separate audio file can be played alongside the video, other changes to default language can also be made in the CPL without changing the contents of the IMF package. The benefits of this are described on Netflix Tech Blog:

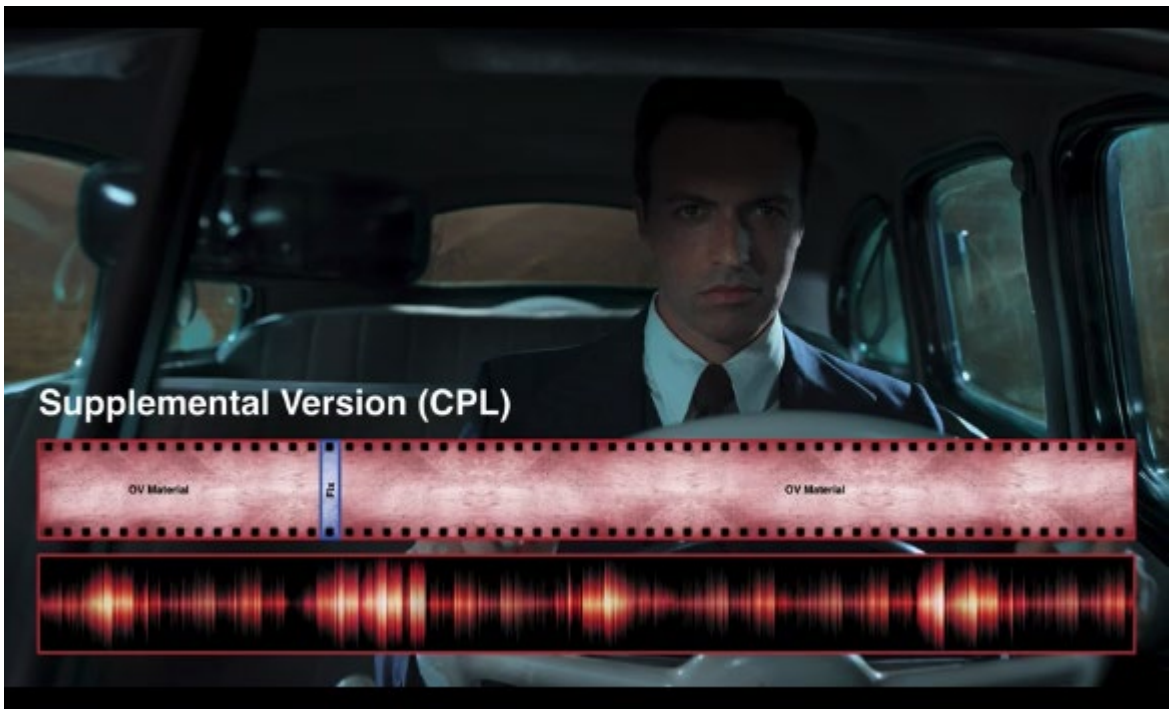
> For a title like Narcos, where the video is largely the same in all territories, we can hold the Primary AV and the specific frames that are different for, say, the Japanese title sequence version. This reduces duplication of assets that are 95% the same and allows us to hold that 95% once and piece it to the 5% differences needed for a specific use case.
>
> (Fetner and Kensworthy, 2016)

The system allows for other reported efficiencies in post production. For example, if any small faults are found in quality control relating to specific frames of a programme, the content producer can fix those frames and only send Netflix the video file with those frames in an updated CPL. Based on timecodes, the updated CPL will then be able to swap-in the corrected frames before swapping back to the main video again. The impact of this is lower file transfer costs and faster updates to content. IMF has added flexibility because it accepts plug-ins. The use of the format allowed Netflix to switch to Timed text in IMF in 2019, providing efficiencies in data storage and delivery, and allowing it to be reportedly more responsive to accessibility enhancement requests in the UK, alongside other streaming giants.

Seeking to share its knowledge and influence the creation of standards in OBM, Netflix has reportedly funded a number of open source initiatives related to IMF.



Original US English version and Chinese version with illustrated content playlist (Source: Netflix/YouTube 2018)



To fix a pixel glitch only requires the fix to be supplied with an updated CPL (Source: Netflix/YouTube, 2018)

Content producers must provide all television and film content to the platform in IMF. When Netflix accepts an IMF package it checks that the sources delivered are 'pristine and guaranteed to be usable by *the* distributed, cloud-scalable Netflix trans-coding engine'. In other words, it is the job of content producers to ensure that the content is object based and in a format that is efficiently planned for cloud storage and multinational distribution.

AMWA AS-11: digital standard used by UK broadcasters

In contrast, major UK broadcasters such as the BBC, BT, C4, and Sky require delivery of air-ready files that comply with AMWA AS-11 specifications and standards developed by members of the Digital Production Partnership (DPP). The UK standard for air-ready masters, AS-11 essentially comprises a pre-mastered audio-video file with minimal editorial and technical metadata to manage aspects of delivery. Re-editing of files after delivery requires the entire file to be resupplied as a new version. Re-versioning requires multiple versions of the same file to be produced, supplied and stored by the broadcaster. Large amounts of metadata that could enhance innovations in consumer-side offerings (such as Amazon's X-Ray service) are lost through delivery to broadcasters of a 'flattened' file. Basic static metadata is required by many broadcaster's AS-11 specification, but many complain that only the bare minimum of metadata is usually produced or delivered.

AS-11 and IMF: not either/or

IMF and AS-11 are not competing formats. As both the DPP (de Pomerai, 2019) and Netflix explain, IMF can be thought of as a pre-rendering format for AS-11. Whilst it is possible to render an IMF package in the cloud (see Rendering in Trends section) it is just as feasible to use IMF to author multiple downstream formats such as AS-11. The crucial difference for broadcasters is that an IMF based workflow provides greater efficiency in terms of file storage and transportation and more flexibility in terms of personalisation and reversioning. Problems highlighted in OBM workflows regarding the loss of metadata could be solved through a more widespread adoption of IMF or a similar interoperable format. By requiring content providers to supply more media objects and more metadata, Netflix and other large OTT broadcasters have more flexibility to automate post-production and distribution and improve standards in their content over time, resulting in a higher quality experience for the consumer. It also mitigates the need to recover metadata through costly AI or human post-production processes.

Other area-specific open standards for metadata management exist such as SportsML for broadcasting sporting events. However, there is not generally enough incentive to adhere to these standards. Whilst there is take up within some broadcast organisations, others take a more *ad hoc* approach based on internal requirements.

Open Standards for Next Generation Audio such as the Audio Definition Model for file-based production (ADM) and Serial ADM for live production are becoming well supported and are covered below.

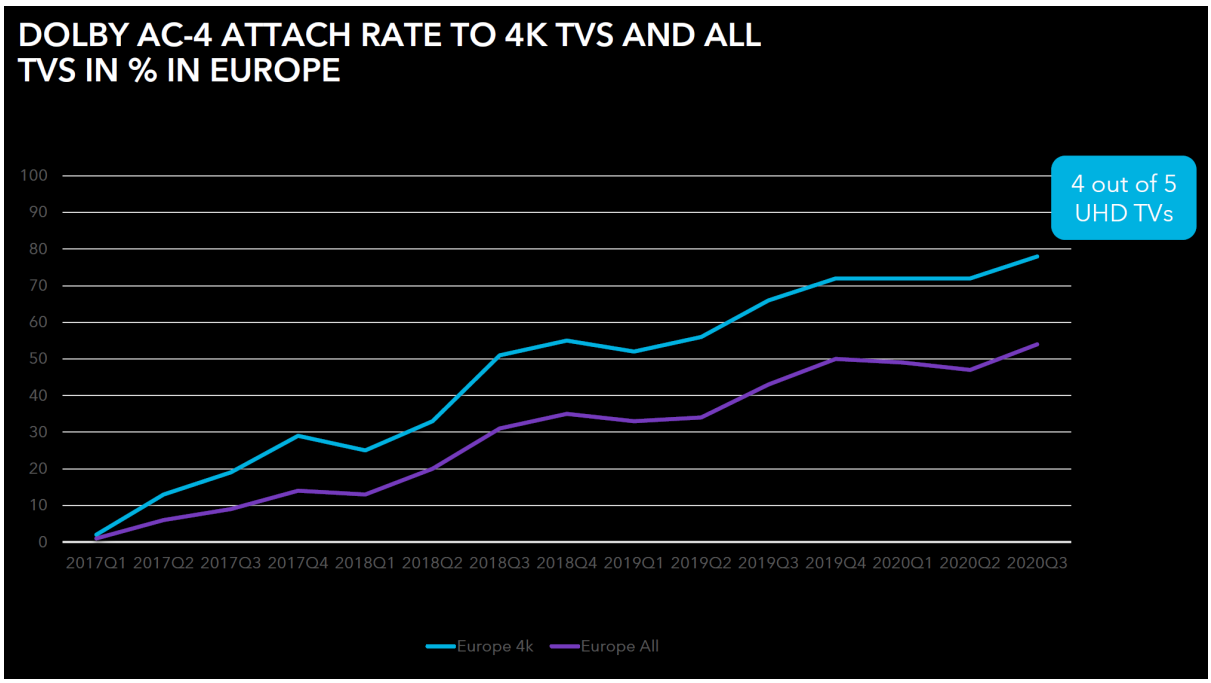
Next Generation Audio

Next Generation Audio, which allows the combination of Objects, Channels, and other audio formats such as ambisonics and binaural audio to be carried, is one of the recent success stories of the Object-based approach. Next Generation Audio is now present in at least 4 out of 5 new UHD TV sets in Europe and the UK, and most new mobile phone models. It is also included in the national and international TV specifications used in many European countries such as the Italian UHD book, Nordig 3.1, FAVN French specification, Polish TV specification, as well as worldwide in South Korea, USA, and the new systems considered in Japan, Brazil

and South America. International standardisation is also mature in this area with DVB, HbbTV and ATSC3.0 having already provided NGA specifications. Codecs for the carriage of Next Gen Audio include MPEG-H, Dolby AC-4 and DTS-X.

Commercial broadcast services using NGA are already on-air in South Korea (MPEG-H), North America (AC-4), and more recently for some events in Poland (AC-4). Immersive audio services are widely adopted by most leading streaming and pay TV platforms. Music platforms worldwide are also starting to deliver immersive music (DTS-X, MPEG-H, Dolby Atmos/AC-4).

Although not yet widely used, the capabilities of these new codecs bode well for object audio use in the near future and are likely to increase the use of advanced accessibility, personalization and immersive audio.

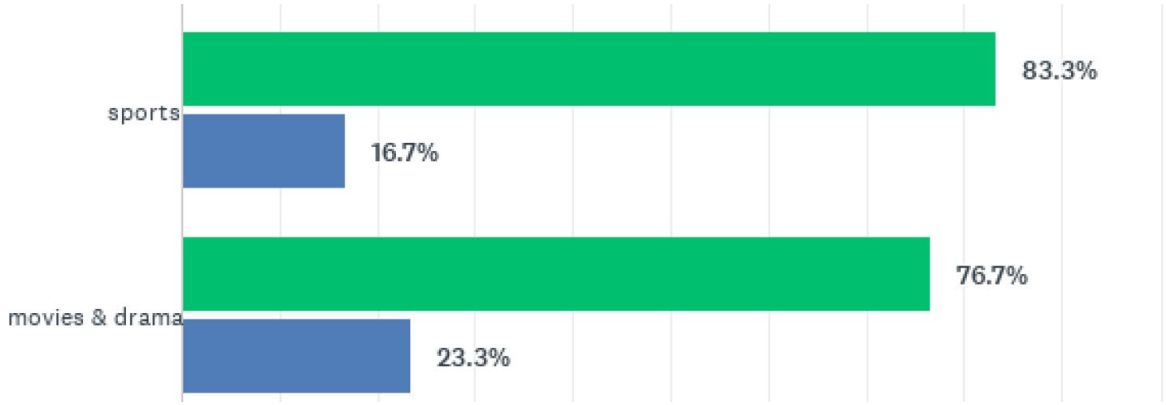


30m TVs >680 models shipped in Europe with NGA - (Source Dolby)

The following benefits of NGA are cited in a recent DTG members-only report (DTG, 2020) in October 2020 as being the most important:

- Immersion: creating an experience that fills the senses via any device, from headphones to home cinemas.
- Intelligibility: for accessibility, enhancing the dialogue over the background or adding descriptive audio.
- Interaction: the ability to choose different languages, adjust the balance, select different commentaries.
- Efficiency: a single asset or stream will deliver the optimum possible experience on any device, from earbuds with a mobile phone to a full home cinema setup.

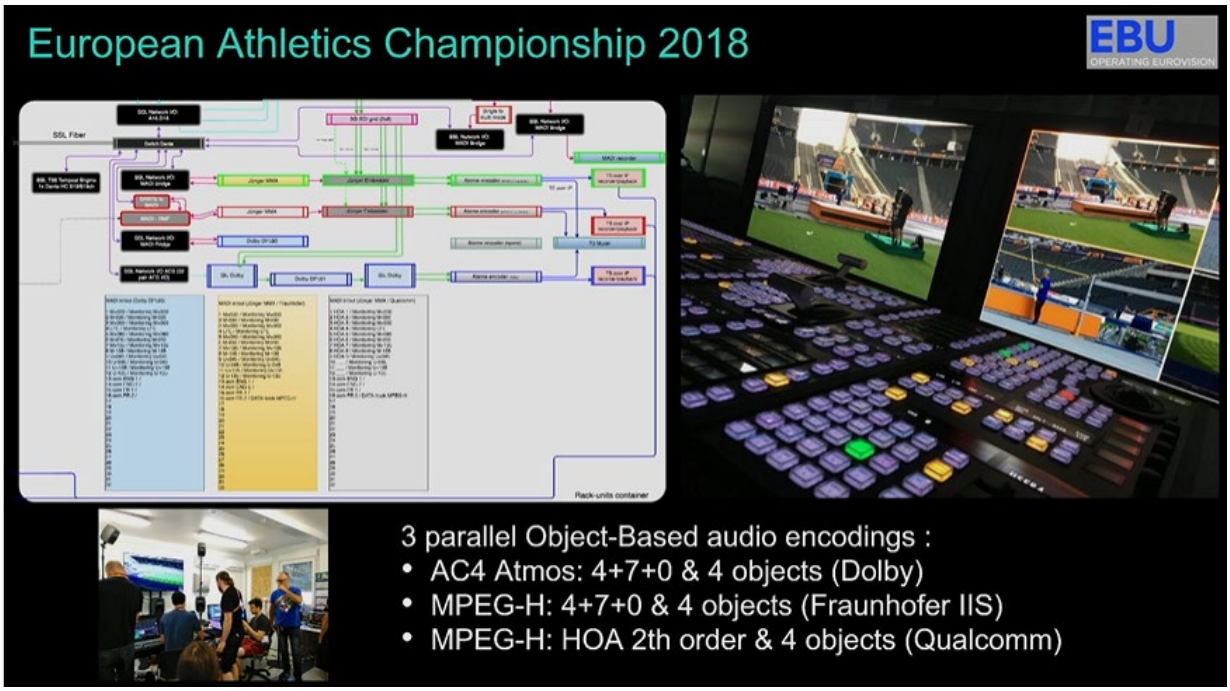
The same report, based around an industry survey, revealed that over 75% of respondents thought NGA to be a mass market feature for sports, movies and drama (with the remainder considering it to be a “nice to have”).



DTG survey asks if NGA will be a mass market feature GREEN or a “nice to have” BLUE for certain genres (source: DTG, 2020)

Regular commercial broadcast services using NGA are already on air in Poland (AC-4), South Korea (MPEG-H) and North America (AC-4), and immersive audio services are widely adopted by most leading streaming and pay TV platforms. Music platforms worldwide are also starting to deliver immersive music (Dolby Atmos/AC-4, DTS-X, MPEG-H).

Although the use of discrete audio objects in NGA services has so far been limited to alternative dialogues and dialogue enhancement as described in this press release from Fraunhofer on the launch of the regular service in South Korea https://www.iis.fraunhofer.de/en/pr/2018/20180614_AME_Worldcup.html, further expansion of this pallet is set to continue over the next few years. One of the most ambitious trials in NGA conducted so far was at the European Championships in 2018 where many of the techniques mentioned above were trialled effectively in a live production environment.



Example: [Demonstration by Fraunhofer at the Eurovision Song Contest in 2019]

This demonstration of MPEG-H at the 2019 Eurovision Song Contest shows the consumer-side benefits of NGA, including multi-language commentary and dialogue enhancement.

(Source: Fraunhofer, 2019)

NGA driving UHD standardisation

NGA is being seen as an important part of UHD standardization. The most recent of these standards efforts and perhaps the most ambitious so far is the TV 3.0 standardisation effort in Brazil. It includes alternative and layered objects as well as the ability to modify audio object positions and loudness.

use case		minimum technical specification			over-the-air delivery	Internet delivery
AC1	Enable immersive (3D) audio.	AC1.1.1	channel-based	2.0	required	required
		AC1.1.2		5.1	required	required
		AC1.1.3		5.1 + 4H	required	required
		AC1.2	object-based		required	required
		AC1.3	scene-based (HOA)		desirable	desirable
AC2	Enable end-user interactivity/personalization when allowed by the broadcaster (e.g. switch among different languages, sports commentators, adjust the commentator loudness level and position).	AC2.1	switch components (audio objects and alternative full mix substreams)		required	required
		AC2.2	adjust object loudness		required	required
		AC2.3	adjust object position		required	required
		AC2.4	enable interactivity when using external sound reproduction devices		required	required
AC3	Enable audio description delivery in the same stream as the main audio, as an alternative full mix or as an additional audio object with associated metadata.	AC3.1	audio description delivery in the same stream as the main audio		required	required
		AC3.2	audio description delivery as an alternative full mix		required	required
		AC3.3	audio description delivery as an additional audio object with associated metadata		required	required
AC4	Enable emergency warning information delivery using audio description.	AC4.1	emergency warning information audio description		desirable	desirable
AC5	Enable a single delivery format for multiple audio playback configurations (TV loudspeakers, soundbars, home theaters, binaural).	AC5.1	flexible loudspeaker configuration render		required	required
		AC5.2	binaural render		required	required

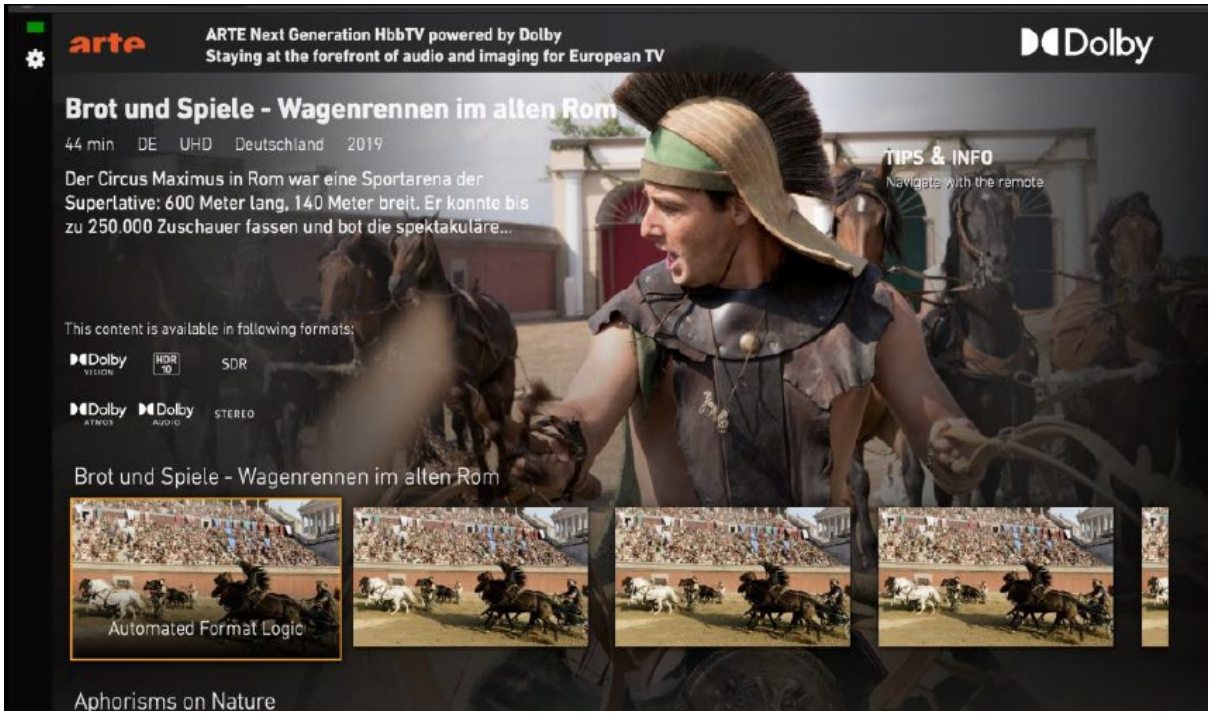
Extract of the Brazil TV 3.0 RfP showing an ambitious request for object-based audio use-cases as standard.

Many other European, country standards and international technology standards already adopt NGA, including: (This is not an exhaustive list and is an evolving picture)

- International standards
  - DVB
  - ETSI
  - HbbTV
  - ATSC
- Europe
  - Italian UHDFI
  - NorDig Standard (Scandinavia, Ireland, Iceland)
  - FAVN France
  - Polish TV standard (On Air)
  - Freesat UK
- Worldwide
  - NABA USA/Canada
  - TTA South Korea (On Air)
  - SBTVD Brazil

Many other countries are in advanced NGA trials and are likely to finish their standardization in the coming year and many, if not most, pay TV and OTT operators in the UK and globally are already including NGA in their current or future services.

By using standards such as HbbTV 2.0.2 broadcasters can offer a level of NGA that was only previously possible via proprietary native OTT apps (e.g. Netflix and Disney+.)



NGA utilised over HbbTV. - Source: Dolby

ARTE created a pilot platform with a new interactive service based on the latest approved 2.0.2 HbbTV standard, supporting UHD in HDR as well as next generation audio. Viewers are able to access TV content in Dolby Atmos immersive audio (via the AC-4 codec) from a standard HbbTV-enabled television.

NGA utilised over HbbTV (Source: Fraunhofer IIS)

At IBC 2018, Fraunhofer presented a Demonstration of MPEG-H in Vewd-powered HbbTV decoder, allowing users to select between different versions of the content such as “default mix” or “dialogue enhancement”, as well as to choose between various available languages. Additionally, for content with accessibility features, the user can select the audio description in their preferred language.

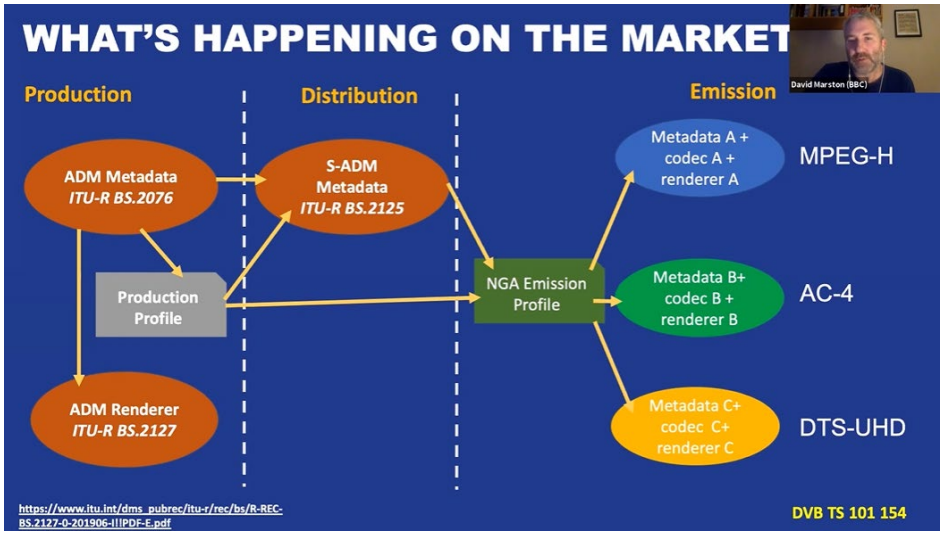
(Also see NGA audio rendering [p23]).

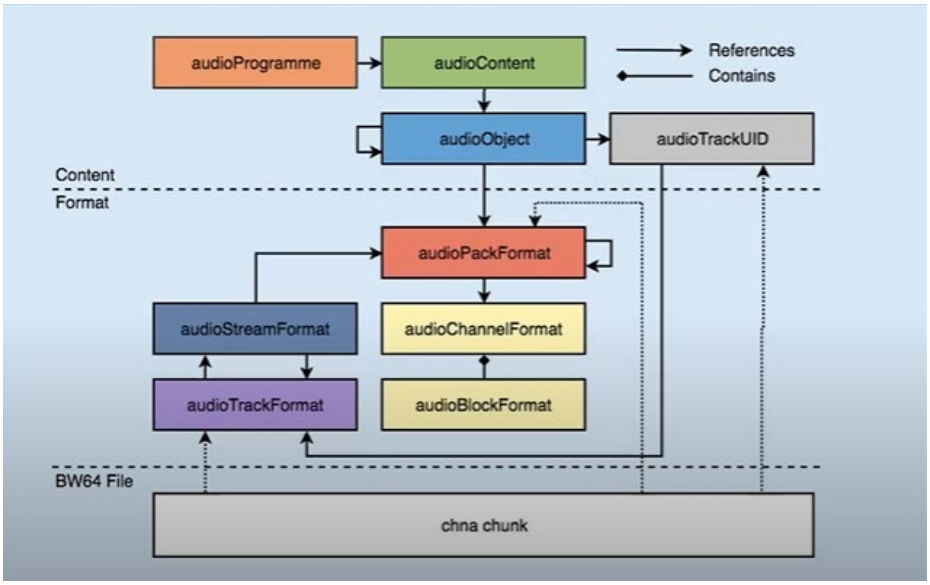
EBU Audio Definition Model (ADM) Renderer

Similar to IMF for general audio-video content (see open standards for OBM), the EBU's ADM and ADM renderer supports a platform-independent format for the creation and exchange of NGA content. Like IMF, ADM is relatively output agnostic and can be used to create compatible content for all major distribution ecosystems such as Fraunhofer MPEG-H and Dolby Atmos.

ADM metadata allows broadcasters to set interactivity bounds on objects as well as defining which objects can be selected together. This information accompanies the audio and is used by the renderer to recreate a personalised audio version.

Both ADM metadata and the renderer are being standardised at the ITU with support from the EBU and other audio industry partners.





The ADM file format as presented by Dave Marston as part of EBU event (source: Marston, 2020)

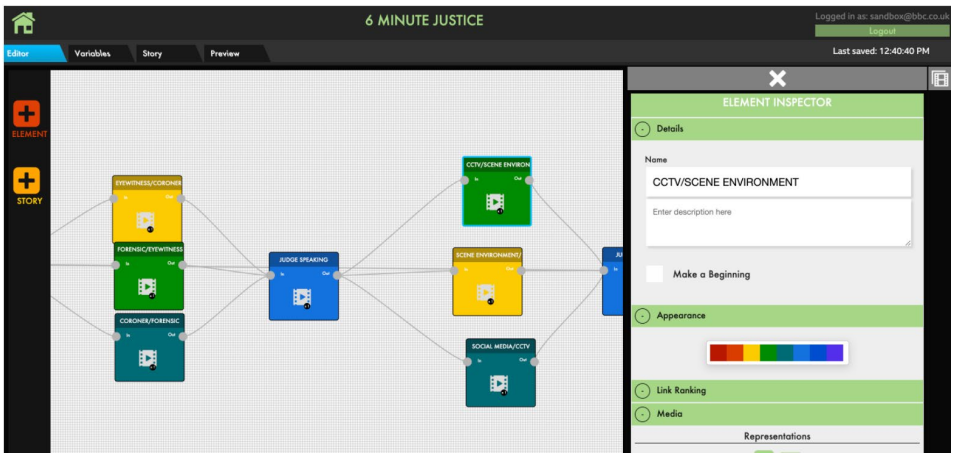
NGA Software and hardware availability

A wide range of Next Generation Audio software and hardware tools are available from a number of suppliers. These tools allow authoring of NGA for both live and post-produced content as well as conversation from open formats such as ADM to emission formats, (e.g. MPEG-H). Quality control software is also available, and it is possible to extract speech from legacy formats (stereo or 5.1) and create from it NGA audio with improved dialogue audibility.

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

StoryFormer production tool



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

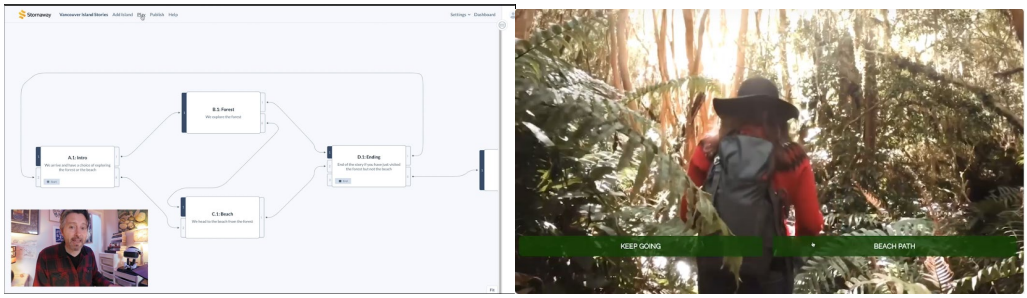
StoryFormer is a BBC R&D tool for creating flexible, responsive stories. It uses branching narratives in its editing user interface, to allow producers of content to create complex relationships between audio and video clips without advanced technical knowledge. The tool and its makers offer an expanded 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 Storybox 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 non-commercial usage. Other tools using object-based production methods within Storybox 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.

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



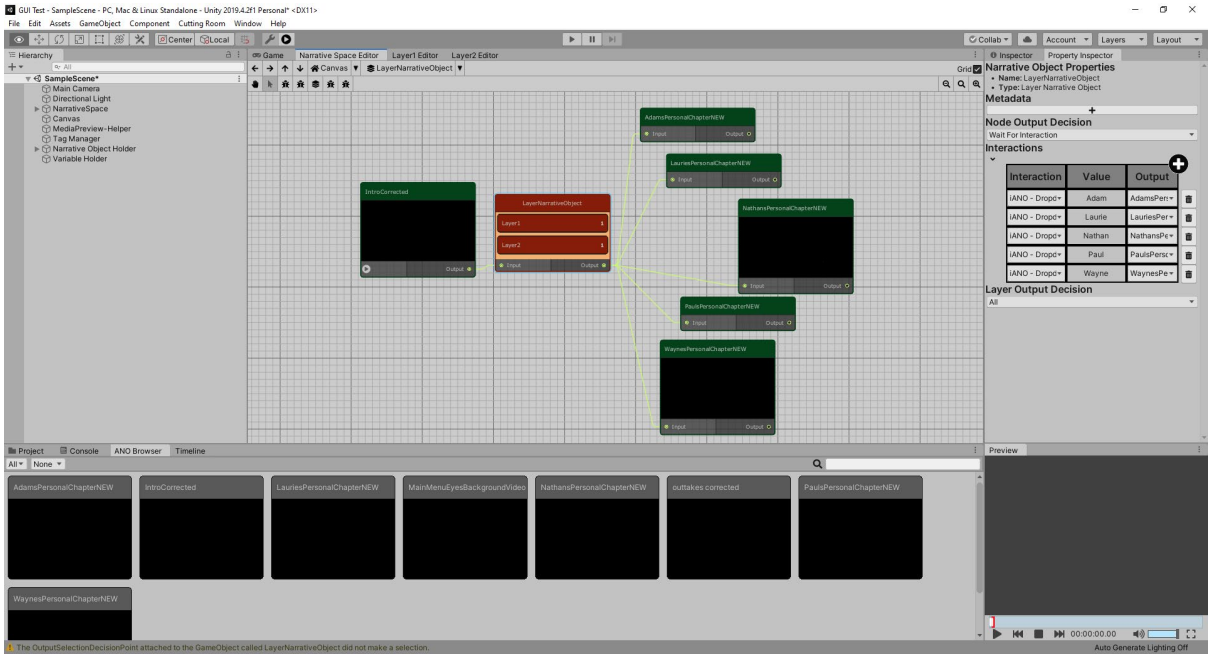
Commercial services such as Stornaway.io allow branching narrative video to be easily created (source: Stornoway.io, 2021)

Stornaway reports 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.

A similar tool by the University of York called Cutting Room (Ursu et al., 2020) was developed in partnership with BBC R&D and funded by the EPRC. A more advanced version

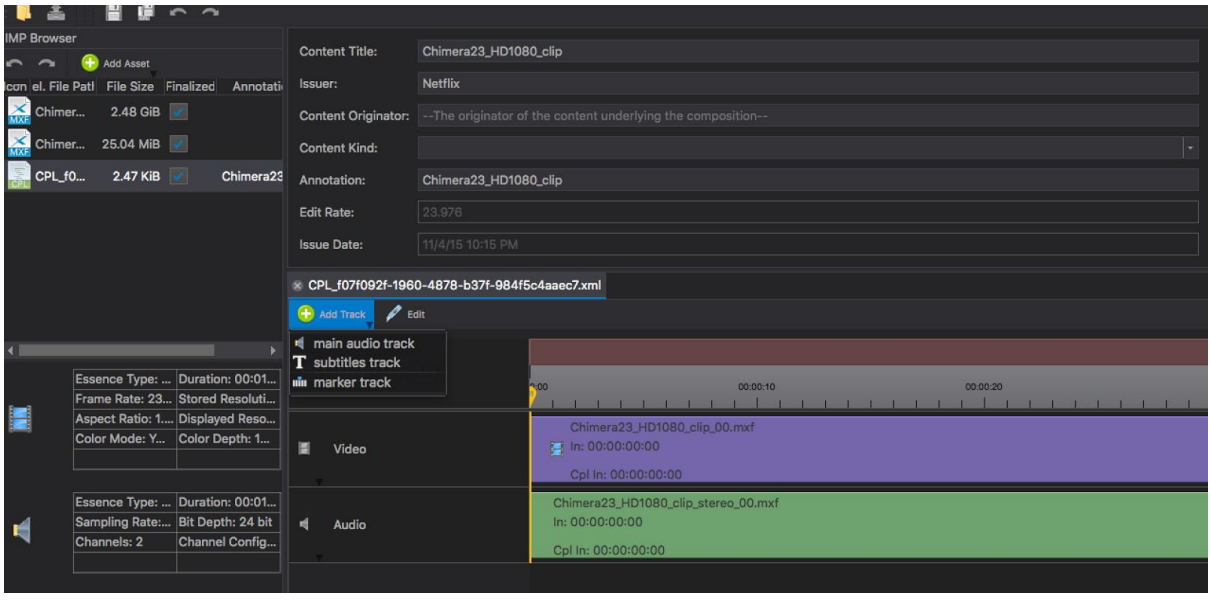
is now being deployed in several University of York XR Stories projects, including an interactive podcast and an interactive video narration for the National Media Museum in Bradford. Cutting Room allows for the creation of both ‘atomic narrative objects’ that can be laid in chunks in a branching narrative 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: Ursu et al. ,2020)

Working with layers: 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 Existing standards for OBM.)



Screenshot from IMF CPL Editor (Source: Ruppel, 2020)

The tool is open source with 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 (DSR Corporation, 2021), which creates 'flat' video files such as DPP compliant AS-11.

Adaptive rendering and edge computing

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 upcoming 5G rollout will improve penetration of broadband across the UK (Ofcom, 2020) and will also mean that for the first time the traditional DTT/satellite broadcast path to most people's homes (which in the UK for the seven terrestrial multiplexes totals 206 mbps total capacity) will be less than the OTT/broadband path through their phone or other 5G device (which can deliver speeds of over 700 mbps, and averages of 100-200mbps). This will create stable conditions for the growth of more adaptive rendering via edge computing.

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 is 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" (Lomas and Lumb, 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.

6. Consumer trends

National and international events

Sporting events with less technological fanfare

Upcoming national and international sporting events are expected to drive adoption of object-based media practices as consumers seek increasingly personalised experiences in their media consumption habits (PwC 2019.3) and broadcasters seek to gain lucrative market share. The technical case for object-based media live transmission was tested during the 2018 FA Cup Final through the 2Immerse project. Today, services such as BT Sport Player (BT, 2020) offer viewers the opportunity to playback key moments of a large number of games and sporting events from multiple angles alongside other augmentations based on objects, such as match events with timeline based navigation. Whilst many international events have been postponed since March 2020 due to the Coronavirus Pandemic, the expected restarting of international competitions in 2021, including the postponed Olympic Games and the Football World Cup are expected to grow demand for OBM both in the UK and globally.

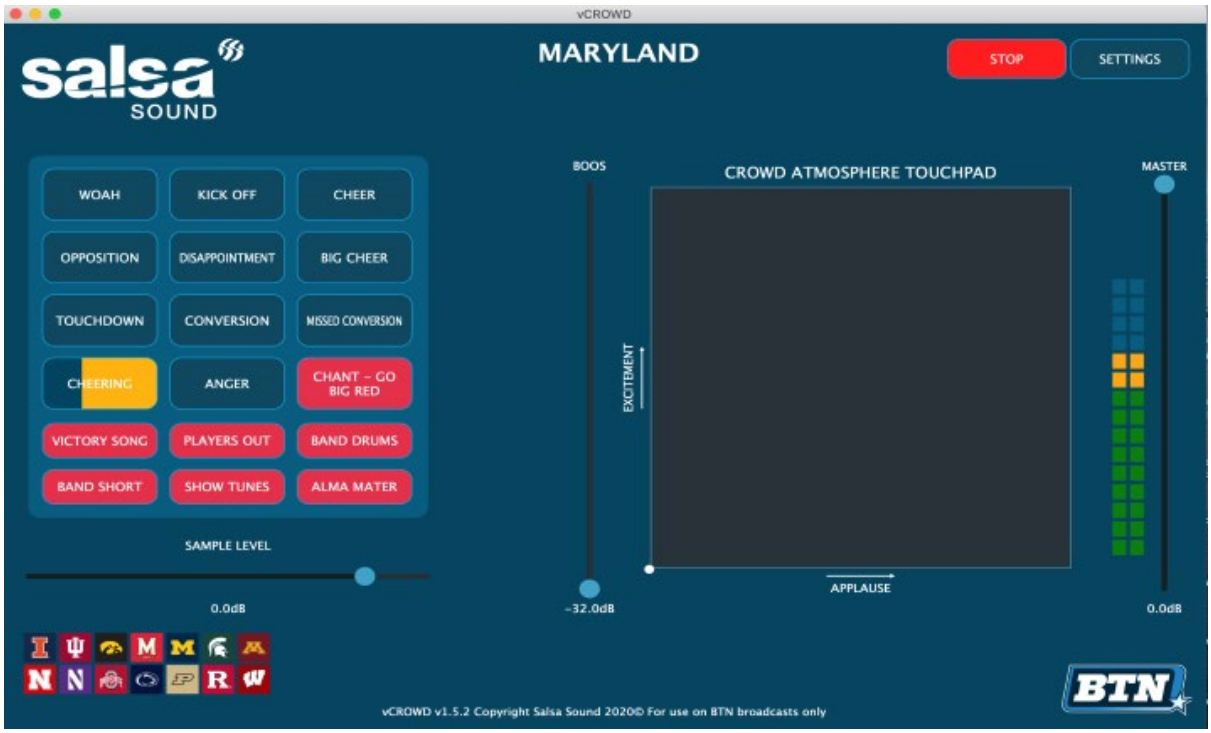
For the Tokyo Olympics, VR and other immersive experiences have been planned by technology sponsor Intel (IOC, 2019) to allow fans who are unable to travel to Tokyo due to Coronavirus restrictions to more deeply engage in a wide range of sports and events, including the Opening and Closing Ceremonies, athletics, gymnastics, boxing and beach volleyball. These experiences will rely heavily on object-based media such as spatialised sound and 360 videos. However, it is important to note that currently many of the usual trials and research projects that use these major events to power their experiments and trial services have been postponed due to Covid-19. Our interviews for this report suggest that very few, if any, major OBM trials are planned around sporting events in 2021.

One exception is the large request for proposal and major test events surrounding the definition of the TV 3.0 platform (SBTVD, 2021) in Brazil, which is designed to become the TV platform for South America in the next ten years. Their requirements are very ambitious in both audio and video capabilities and include Object-based use-cases. The platform will be IP based for both broadcast and OTT.

Coronavirus pandemic

There is now clear evidence that the event of the Covid-19 pandemic itself is speeding up changes in consumer media habits (Balhaus and Chow, 2020). Lockdowns in the UK and globally are accelerating the uptake of digital media services as people seek to find new forms of entertainment to cope with more isolated lifestyles. (Disney+ powered its way to 100m subscribers in record time during the pandemic). The switch to on-demand broadcast models offered by both UK and US providers is changing the competitive landscape. On these channels, consumers are adapting to new standards, powered by object-based media, such as universal multiplatform availability of accessibility features - audio description and subtitles - as well as other enhancements, such as immersive object-based audio as standard (e.g. Dolby Atmos and DTS-X).

In addition to the uptake of OTT, Covid-19 conditions have also forced broadcasters to find novel object-based 'virtual' production methods to enhance live programming. At the time of writing no crowds are allowed to gather at football matches and other sporting events. To solve the problem, UK company Salsa Sound (Machine Learning, p27) used its object-based collection of spatialised sound to provide realistic crowd sounds. Its new commercial product vCROWD allows the real-time creation of a virtual crowd sound to perfectly match the action on the pitch. An operator watching the match live controls crowd sound using a touch screen interface and a trackpad which allows them to vary the mood of the virtual crowd, controlling excitement, applause and even boos and whistles. All sounds are from genuine home crowd recordings meaning that the songs and chants of the crowd are authentic to that stadium. In the US, isolated crowd sound for NBA games was made possible by using samples from the NBA 2K computer game. Whilst crowd noise issues may soon be mitigated by the return of spectators to spectator sport, these examples point towards the adaptability of object-based production approaches and also the convergence of object-based broadcasting production and game production methods (Game Engines p29).

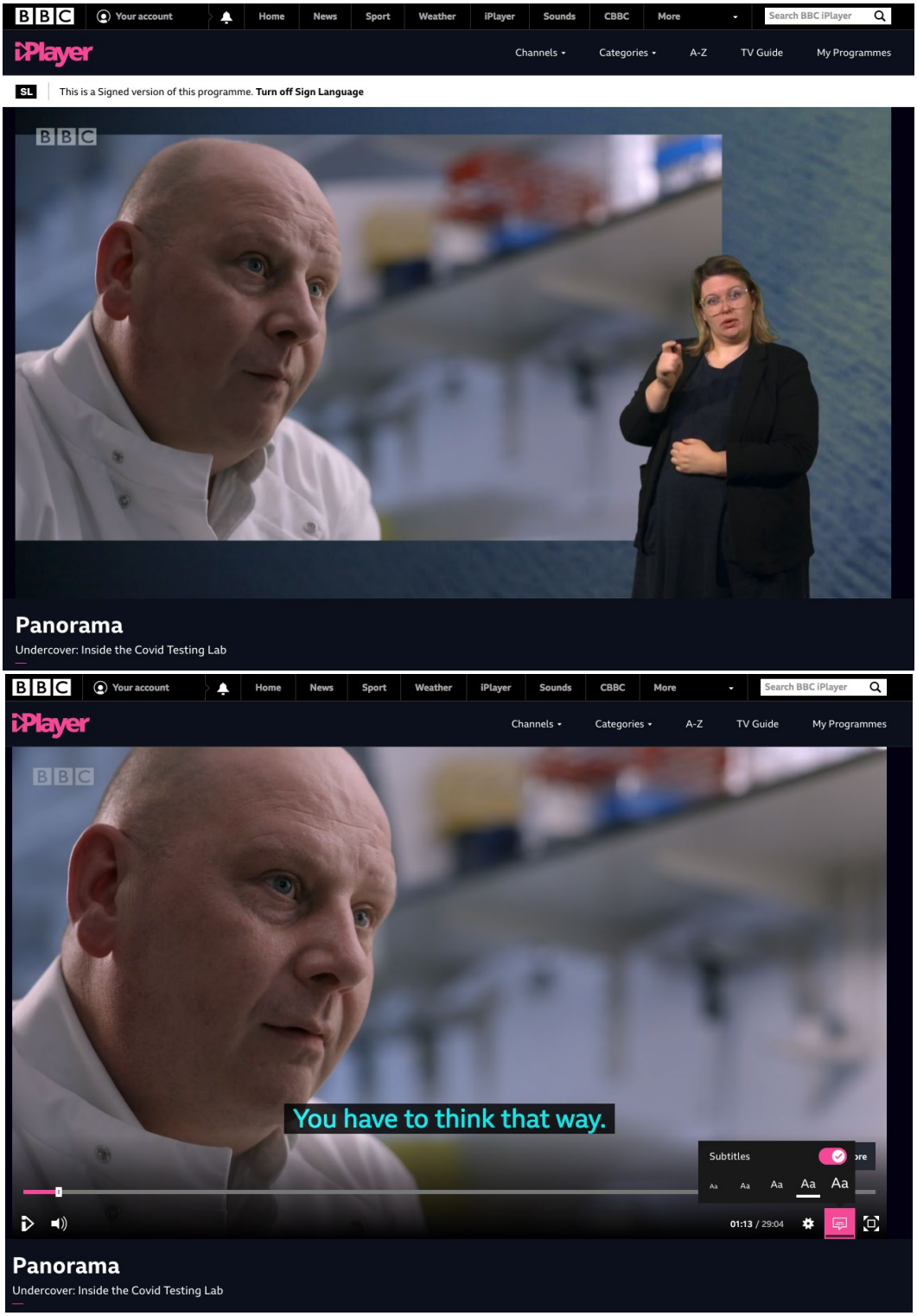


Screenshot of the vCrowd interface (source: SVG Europe, 2020)

Increasing demand for accessibility

The most pressing and perhaps most prevalent use of personalisation made possible by objects is extending accessibility. Services like Voice Guidance offered by some television manufacturers show how existing metadata provided as part of a standard AS-11 output, such as programme title and EPG, can be adapted by manufacturers to provide spoken word guidance to blind and partially sighted audiences (RNIB, 2020). Digital-only and hybrid devices also allow for the delivery of object-based services and versions of content. For example, BBC iPlayer offers a wide range of accessible services including AD, signing and subtitles for a

large percentage of its recorded programming, with OBM features such as text size manipulation made possible through a dynamic text layer object rendered on top of the standard video.



iPlayer provides signing and subtitling options for a percentage of its programming (Image source: BBC iPlayer, 2021)

However, UK major broadcasters have a public service mandate to provide programmes to a wide range of legacy devices including traditional television sets as well as a commercial drive to share content across a plethora of third-party services. Both of these factors make standardisation of these features increasingly difficult. Where media providers are metadata rich throughout their production and consumption workflows and have control of rendering on proprietary platforms (Netflix, Amazon Video and Disney+, for example), consumers are responding positively to the reliability of accessible features for these services.

So, accessibility features are now becoming a feature for all, not only those with specific needs. But only the giants are able to deliver these services with consistency and reliability across all platforms and devices. Accessibility groups are consistently seeing that OTT giants are treating accessibility as a standard feature (not a chore and a burden) and are ensuring that it is available and consistently offered across their platform and on all devices. For those who rely on these services, this has been very warmly welcomed and this consistency and ubiquity of accessibility features are causing users of accessibility features to gravitate towards these platforms. The traditional broadcasters often trail behind in terms of the availability and the consistency of accessibility features for their content, both on broadcast and even more so on less prominent catch-up portals/apps. This is understandable due to the drag of legacy platforms, the plethora of delivery methods and the multitude of devices and apps that they support, compared to the carefully curated own-app approach of the OTT giants.

The lack of a holistic object-based approach, standards, storage and cataloguing in production is also slowing things down, leading to the re-authoring of subtitles, AD and other features multiple times for the same content, and only sparse use of the production metadata (scripts/timings etc.) to aid the creation of these features.

Accessibility “objects” are often not part of the production deliverables and are generated by the broadcaster after the programme is delivered. This means the production company may not be responsible for them and if the content is re-sold to others, these objects will be missing. With the increase in the use of AI to produce the first “rough cut” of the subtitling, some interviewees reported difficulties with the background audio masking the dialogue and producing poor AI results.

This has led to experiments where only the dialogue tracks (dialogue objects) are fed to the AI algorithm, improving the transcript considerably. A fully object-based approach would allow the script, clean dialogue, and other elements to be also used by humans or AI to improve the subtitle creation.

Next generation accessibility?

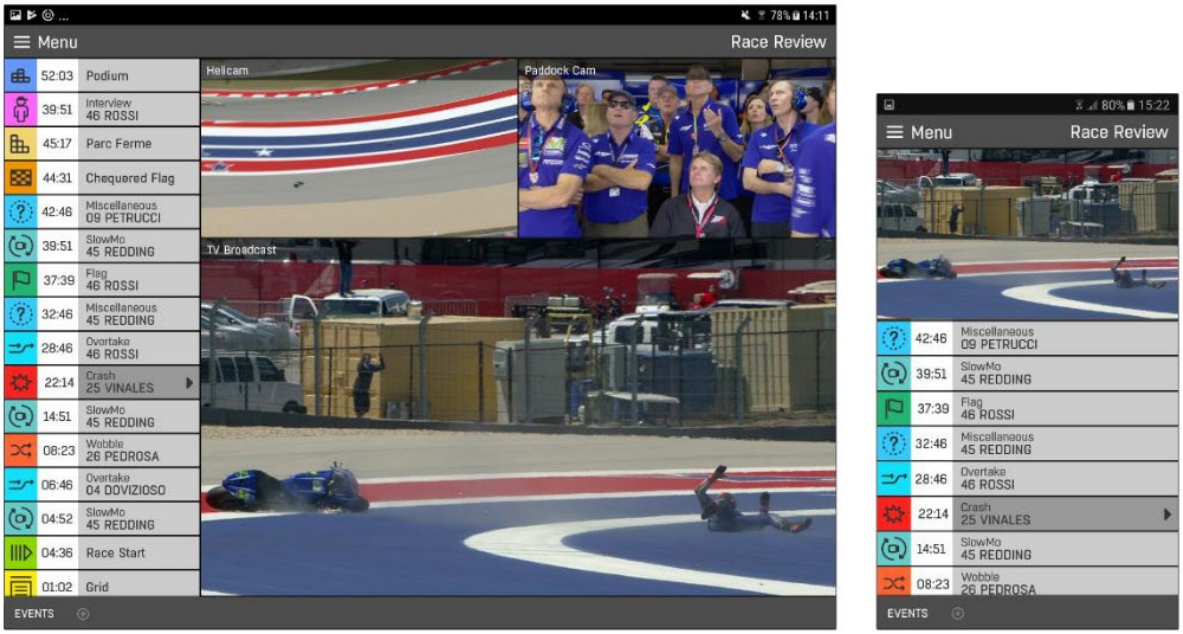
The optionality provided by object-based accessibility also provides new opportunities to respond creatively to forms of accessibility such as AD. A project by the University of York and Anglia Ruskin University, funded by the AHRC, added an Enhanced Audio Description (EAD) object to the traditional AD object to offer new audio description approaches to the narrative. Ideas trialled in the project included first person narration from the perspective of a particular character and the sound bed manipulated with sound effects and spatialisation to provide a better sense of location and facilitate a more compelling experience for blind and partially sighted audience members. Research showed that this provided a more enjoyable and informative experience than AD today (López et al., 2020). The short film Pearl (Palumbo and

Feng, 2015), with an EAD soundtrack by Mariana López and Gavin Kearney, demonstrates approaches to EAD trialled in the project.

Use of multiple platforms and devices

Content destined for one screen and one platform is a thing of the past, as all content is prepared in such a way that it can be consumed on any screen and on multiple platforms. This can result in compromises such as a plethora of versions, or editorial or technical changes in the content or format to cater for the highest and lowest common denominators. Objects could aid this situation enormously, enabling multiple versions to be created from a single object repository as and when needed. Conditional rendering can also allow content to be adapted to provide the best experience tailored for the consumer's device.

An example of this can be seen in the 2Immerse project that developed object-based versions of MotoGP racing footage for different devices with specially created graphical overlays showing rider position times and key elements of information and different video feeds.

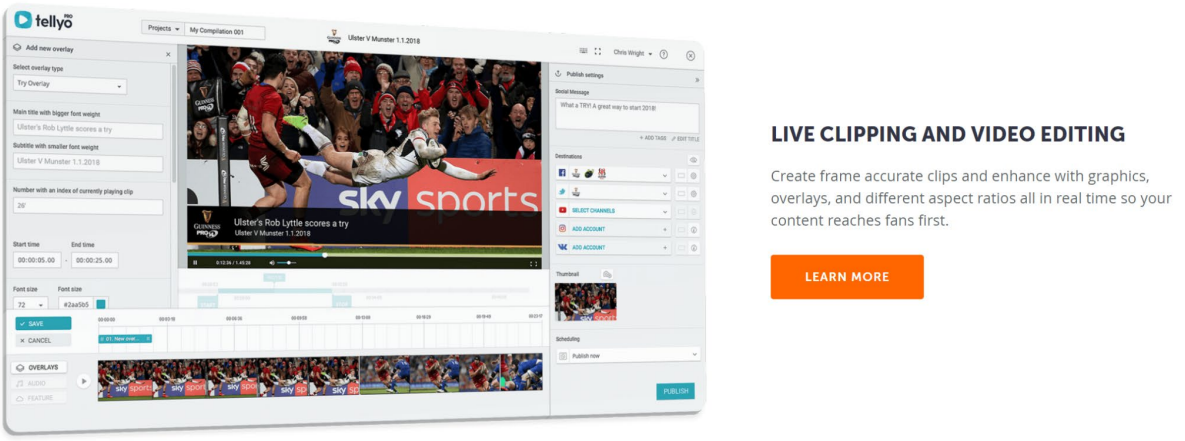


The 2-Immerse project also highlighted the use of graphics to maximise interactivity of on-screen real estate for mobile and hybrid devices. Examples included an interactive score clock and general user interface, providing standard match overview information but also access to stats, team line ups and even picture-in-picture replays. Many of these features are now offered as commercial USPs by 2-Immerse project partner BT and others (see BT Sport Player example in common types of object).

Use of the score clock as an interactive GUI (Source: Kegel et al, 2018)

Shareability, snackability and social comment

Social sharing of short-form content is one of the most powerful forms of content distribution in the 2020s, evidenced by the proliferation of platforms such as Instagram, Snapchat and TikTok which specialise in short-form content, as well as to a lesser extent YouTube. There is evidence that the pandemic is increasing the popularity and value of short form formats (Balhaus and Chow, 2020). Small snippets (or objects) of content, often chunks from long-form content, are shared millions of times per minutes, via public and private groups, on websites, in tweets and posts. Content that is not shareable is likely to miss out on ratings and viewership. Even today, however, content providers are not making it easy to share content objects. Media is often crudely captured and clipped illegally with reduced quality and inappropriate formatting. Objects and metadata may well improve this situation and ensure that viewers can tweet and post a link to a particular object or objects from within the content, giving the content provider the ability to see the popularity of these snippets directly via the traffic they create in the same way as web hits and post views are captured today. Content made for “Social” may well embrace objects to aid shareability and tracking. Companies such as Tellyo are helping operators to bridge this gap.

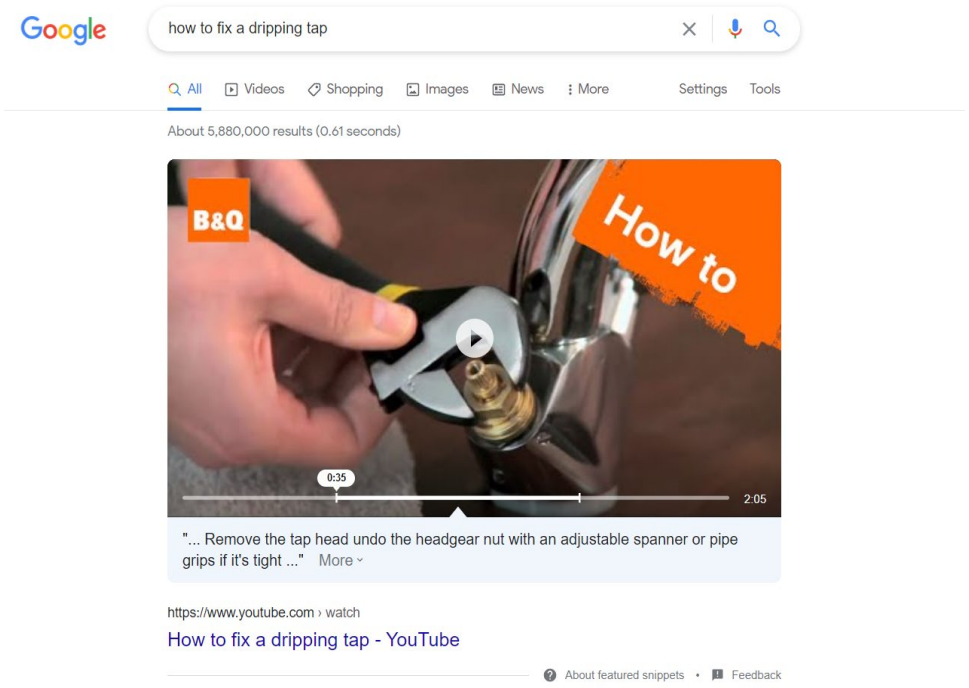


Attitudes from broadcasters and content providers towards providing objects and metadata that the consumer can reference, comment upon and link to are however still evolving. Many can imagine the benefits of linking to a snippet of a programme in the same way as you post like to web content today.

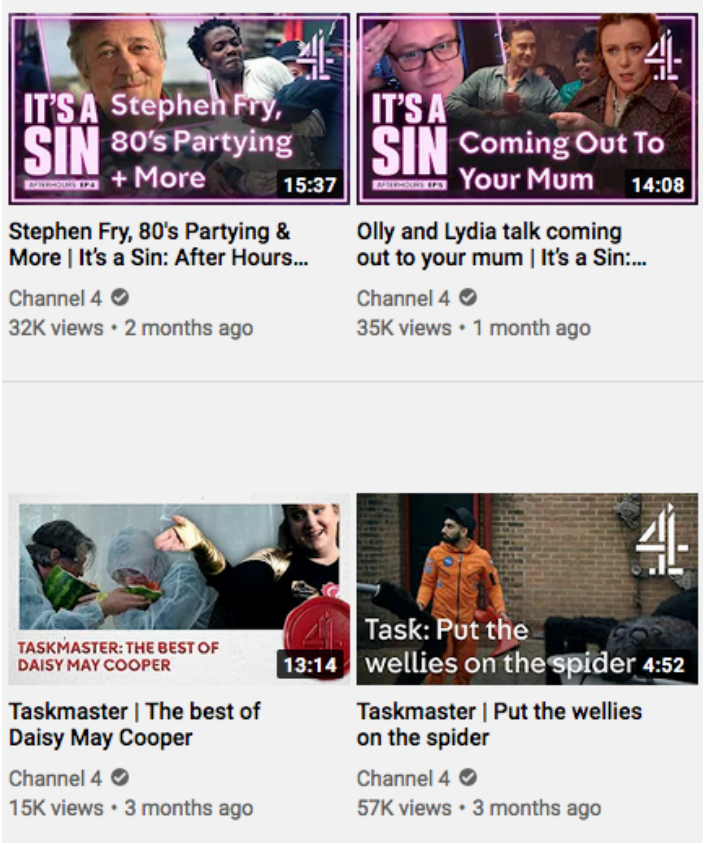
Use cases such as “did you see this on TV just now, **<unique object link>**” is a great concept. However, the practicalities soon become daunting and the level of access to hitherto unreleased metadata worries some. Imagine being able to search, find and play car-chase scenes, or find the appearances of certain actors in a VoD library. A more liberal access to objects might allow a fan to remove the broadcaster’s commentator from a clip of a sports match clip and allow them to add and share their own comments with friends.

Example: YouTube Suggested Clips

YouTube Suggested Clips are a new way to feature part of a video, directly within a set of Google search results. They play back the most relevant section of a video. The user does not need to leave the results page on Google to see this video.



Programme makers themselves are going through their own archives and creating short-form snackable content created from snippets of archive material and putting them out on major social media channels such as YouTube, Facebook and Twitter. For example, Channel 4 has a YouTube channel featuring best bits and scene clips from its drama and comedy programming. This approach benefits from being more easily searchable, shareable and snackable with promotion down the line with links to DRM versions of Channel 4 content.



Channel 4 on YouTube (source: Channel 4/YouTube, 2021)

With well-tagged *chunk* objects the creation process for such content and the ability of consumers to navigate the clips they enjoy could increase dramatically allowing searching terms such as “coming out scene on It’s a Sin” or “Daisy Cooper’s best scenes on Taskmaster” on the fly directly on broadcaster channels.

Voice search and navigation: practical and playable content

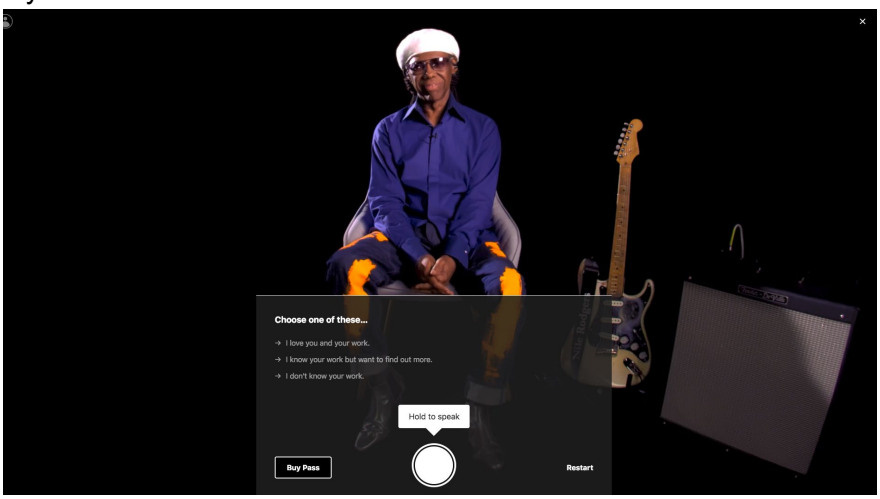
Although browsing through the EPG/channels is still ‘alive and well’, the previous section indicates the strong shift towards search and recommendation. With the number of voice assistant enabled devices predicted to double to over 8 billion worldwide by 2024 (Vailshery, 2021), consumers are expected to increasingly use voice to navigate media content this way. “Hey TV show me Daisy Cooper’s best scenes on Taskmaster” is now a perfectly valid phrase, but also a hard one to respond to. Searching within an hour-long programme with no objects or metadata is very hard. Whereas finding the car-chase scene or a phrase is easy if the content is organised using a metadata-rich format with accessible chunks such as well described scenes, layers or searchable subtitle text. The key to a great experience is therefore the ability to search the metadata of the content available to build a good match.

Voice-based interaction with chunk objects is also forming a key part of exciting lean-forward interactive experiences such as Charisma.AI and broadcaster Sky's work on the voice-based experience for Bulletproof 2. The viewer must help the protagonists of Sky's Bulletproof police drama to interrogate a criminal by speaking with him. The content uses standard chunks of media content and conversational agent objects that allow a conversation with the suspect. These trigger video-based outcomes to provide a high-octane crime drama experience.

Web-based interactive crime drama powered by voice (Image source: Sky Studios,2020)

The use of voice agents has also been used in a project led by Forever Holdings Limited in collaboration with Bright White and Manchester Metropolitan University funded by Innovate UK to power an interview with famous musician Nile Rodgers.

In the Room with Nile Rodgers uses over 350 answers to questions recorded as separate objects to allow users to speak with Nile about his life and career using conversational agent technology. The project developed new techniques around the rapid creation of high-quality metadata rich media objects for use in an object based conversational agent workflow. The experience has been released in association with the Portrait Gallery and Universal Music and points the way toward new forms of media experience for consumers to benefit from in areas of documentary and news.



Web-based conversational experience with Nile Rodgers
(Image source: Forever Holdings Limited, 2021)

7. Commercial trends

Making the business case for object-based media

The business side of object-based media has for many years been troublesome.

Whilst on one side it offers potential savings in the production of content, content reuse and re-versioning, as well as a new, enhanced, personalized and compelling consumer experience, on the other side it has and is also a very “heavy lift” - there is no instruction manual, there are very few tools, and there is a small and fragmented market to which objects can be delivered. All these factors are manageable if the results are well known and OBM can clearly show that the financial/business/strategic goals can be easily met. The ROI scales have however not tipped in its favour yet, and OBM is in many cases still in the experimentation and feasibility phase. Even where OBM has been launched it has not been shown to be directly monetizable, and savings in production are not yet possible due to the extra work involved in re-architecting the workflow.

There are however notable exceptions and the direction of travel is certainly a positive one:

- The BBC has now created tools that allow all its programmes to potentially utilise OBM.
- The NGA formats (Dolby Atmos/AC-4, MPEG-H and DTS-X) have ensured that all premium content such as sports, movies and drama is created in NGA/OBA.
- User interactivity for branching narratives and navigating chunks of media is now built-in to the Netflix player, Sky’s immersive VR platform, and the BBC’s web player.
- The original and still one of the most widely used objects, subtitling, is now available more universally than ever and has been embraced by the streaming community.
- The industry also seems to be encouraging itself, with many awards and accolades presented to content that has innovated with OBM.

Targeted advertising: an OBM commercial success

One other success story is the rise of targeted advertising, where object-based media has directly shown its value. By substituting one media chunk/object (the original ad) with another (the targeted ad), which is tuned to the individual viewer, advertisers can show higher efficacy. This translates directly to financial and business goals and additional revenue that keeps the TV industry thriving. This has made targeted advertising one of the biggest drivers of change and technology adoption in the content workflow in recent years.

Example: **NHS Smoke Free and Sky AdSense case study**

Public Health England ran an NHS Smoke free campaign using the AdSmart targeted advertising to access members of the public who were most likely to be smokers to inform

them of the dangers of smoking, encourage them to stop and signpost support. Research from Sky suggested that the targeted AdSmart campaign resulted in better recall of key messages by that group (Sky, 2021).

Wider commercial costs and considerations

Object-based media is not without commercial costs and complexity especially during the early phases of its adoption and rollout. These include:

- longer production times (due to the novelty of the technology),
- Increased costs for staff, equipment and planning cycles,
- Increased complexity in the workflow,
- Increased complexity/time needed in preparing additional metadata,
- Increased cost in the storage and retrieval of multiple objects,
- Increased cost and complexity of distributing multiple objects to consumers or cloud rendering a version personalised to each viewer:
    - These include consideration around IP infrastructure. How are rights managed for new versions of content assembled from multiple different parts?

To counterbalance these early disincentives there needs to be positive incentives to the distributors and commissioners of content, and to the producers of content. Like many things object-based, media is “front loaded”, meaning that effort at the preparation and production side is needed way before any benefits for the final experience can be realised or monetised.

The following factors have so far, been the strongest catalysts to tempt commercial entities to start producing with objects:

- **Accolades:** content producers thrive on accolades, it drives the demand for their content, provides them with vital marketing opportunities, and helps them galvanise their reputation. Object-based production is seen as a new frontier and those who explore it can expect accolades for doing so.

- **Regulation:** regulatory rules can go a long way to providing the impetus needed to move towards object-based production. Consider a situation where a producer already needs to produce the programme in 1, 2 or 3 national languages, have subtitling, signing, audio description, and a child-friendly edit to satisfy regulatory obligations. These requirements start to close the gap in cost/complexity between the existing and the new object-based methods and, as the technology improves and becomes more mainstream, the balance may tip towards object based being a good investment.

- **Demand:** This has been a tricky topic for object-based media as the value it offers is multi-faceted and unlike ‘high resolution’, ‘flat screen’ or ‘colour TV’ it is not easy to convey to consumers. There have been mixed results in terms of feedback from consumers to the current trials and it is perfectly valid to say that object-based media even works better when the consumer is unaware that it’s being used to curate their experience quietly in the background. Some applications have however shone through, reactions to the branching narrative pilots have been strong and, whilst they divide opinions, have certainly created interest. Immersive and personalised audio

have also seen very positive reactions, especially from sports fans. BT reported a strong increase in demand for its UHD service when it launched Immersive audio, and Salsa sound reported strong demand for its personalised audio presentation solutions, and due to Covid, its tools for creating virtual crowd sounds.

- **Cost saving:** whilst as mentioned above the initial ramp-up and learning curve towards object-based media might need investment in time, effort and cost, the equation soon tips the other way as savings can be made in re-versioning, adapting, creating accessible or foreign language versions, highlights, bumpers and trailers. The trickier problem may be that the value is distributed throughout the chain and is hard to quantify in terms of direct value to one party (such as the studio, distribution team, or commissioning etc.

Given the costs and benefits of using object-based methods are distributed unevenly across all parts of the content ecosystem, there needs to be a balance of incentive and coercion to get the industry to the tipping point where OBM is universally seen as a net positive.

UK is in leading position in object-based media

On the global stage, the UK can be seen as a leader in Object-based media. From its pioneering of subtitling and red button content to its leadership in interactive and immersive experience design, the UK has been at the forefront of this new technology in research development, trials and commercial services.

Some recent object-based media research projects:

- **2-IMMERSE (2015-2018)** - EU-funded UK research project developed a new open-source platform for object-based multi-screen entertainment with partners BBC, BT, Chyron Hego, Cisco, CWI, illuminations and IRT.

- **Project FacinatE (2009-2013)** - European Framework 7 collaborative project to develop technology for panoramic imaging and 3D audio acquisition, delivery and interactive viewing with BBC R&D.

- **Audience with a Hero (2019-2020)** - an Innovate UK Audience of the Future project developed production techniques for speech-based documentary formats.

- **NDS Fresco (Circa 2012)** - researched massive screens in the home made possible by OLED technology and future formats for these screens reliant on OBM.

- **Orpheus (2015-2018) -** EU-funded HORIZON 2020 research project that pioneered the development of an end-to-end object-based media chain for audio content.
- **Enhancing Audio Description (2016-2018)** - Explored how sound design techniques can be used to rethink accessibility to film and television for visually impaired audiences. Research included the application of surround sound rendering, interactive media systems and first-person narration.

- **The Prosperity Partnership (2021-2026)** - UK Research and Innovation (UKRI) and BBC R&D have announced a new five-year partnership with the universities of Surrey and Lancaster to develop and trial new ways to create and deliver object-based media at scale.

Academia:

- **University of York:** XR Stories funded by UKRI researches and develops for the future of immersive and interactive storytelling.
- **University of Portsmouth**: Portsmouth's Centre for Creative and Immersive Extended Reality (CCIXR) supports innovation in the creative and digital technologies of virtual, augmented and extended realities.
- **Royal Holloway and National Film and Television School: StoryFutures Academy**: funded by Audience of the Future grant, is developing creative training and research programmes to ensure the UK creative workforce is 'the most skilled in the world in the use of VR, AR and MR'.
- Manchester Metropolitan University: School of Digital Arts (SODA): SODA is a purpose-built, interdisciplinary school focused on creative media production.
- University of Surrey
- UCL
- University of Lancaster

Commercial companies and start-ups:

- Mirriad
- SalsaSound
- Finecast
- V-Nova: Advanced video compression software
- Sky Immersive
- BT and BT Sport Player
- Endemol Shine UK
- Stornaway
- Forever Holdings Limited

However, whilst the UK has led the way with many of the early experiments, commercial deployments have been slow to emerge and, according to some UK content providers, there is still some way to go.

As established in our section on metadata some reported that production metadata did not have a clear path to be preserved and used in the object delivery workflow, and metadata and objects were often being lost/combined in production. Some cited the lack of specifications or content guidelines for the absence of objects and metadata, whilst others felt that commercial tools were not yet available to harness the power of objects.

There is however strong evidence that the world is looking to the experiments, tools and methods pioneered in the UK to drive the industry forwards, the open source tools created by the BBC are in widespread use in academia and likely in industry too. Global companies such as IBM, Cisco, Dolby, Fraunhofer, and Netflix have all learnt from collaborations with UK projects, staff and productions.

We are also seeing some very interesting start-ups creating very good work in the area, but there is a concern that these companies lack the investment needed to compete with global tools giants who are likely to, one day, sell complete tools back to the UK's early pioneers.

There is also clear evidence that global companies see the UK as a rich hunting ground for talent in the creative industries and especially in cutting edge media research: BBC R&D and the R&D departments of UK universities often find that their staff and students are in high demand by global corporations and often lose staff and talent to these companies as soon as they gain experience. Experience in particular in the hybrid skill sets required for object based media are still in relatively short supply versus demand (Bennett & Murphy, 2020) and global companies are recruiting heavily in these areas.

8. Compliance Regulation and Competition with Object-based media

Current compliance and guidelines that regulate media were designed for a non-object-based market, where a single or multiple linear versions of the content were produced. The advent of Object-based media throws up some interesting challenges and opportunities for broadcasters and regulators.

Audio Loudness

For linear programming, the loudness of programming and advertisements is well regulated and broadcasters follow guidelines such as EBU R 128 recommendation for loudness normalisation and maximum level of audio signals. It is usually applied during audio mixing of television and radio programmes and adopted by broadcasters to measure and control programme loudness using Loudness meters instead of only Peak Meters (PPMs), which were often the norm until 2010s. The switch from audio peak-normalization to loudness normalization was probably the biggest revolution in professional audio in the last decade.

EBU R 128 developed by the PLOUD group recommends normalizing audio at -23 LUFS ±0.5 LU (±1 LU for live programmes), measured with a relative gate at -10 LU. The metering approach can be used with virtually all material.

To make sure meters from different manufacturers provide the same reading, EBU Tech 3341 specifies the 'EBU Mode', which includes a Momentary (400 ms), Short term (3s) and Integrated (from start to stop) meter. Many vendors support 'EBU Mode' in their products.

The advent of object-based audio however allows users to personalise aspects of the audio, add and subtract and manipulate the level of different elements. The order and chunks of a programme can also change, as can the duration. This adds an additional layer of complexity to monitoring loudness.

In 2020, a new sub-group of the PLOUD group was created: PLOUD-NGA, chaired by Andrew Mason from BBC R&D. The goal of PLOUD-NGA is to find a way, or ways, to measure loudness level in Next Generation Audio (NGA) systems.

The NGA codecs themselves also have mechanisms that ensure that the overall loudness is not unduly impaired by personalization and the ADM standard adds metadata for loudness control.

Technology is however evolving faster than the recommendations and early object-based services may well face loudness issues.

Investigating loudness complaints is also likely to become more complex as verifying any content will require close examination of the objects, the metadata, and the personalization settings of the complainant.

Some are of the opinion that once personalization of the content is evoked by the user, the user has made a conscious decision to potentially alter the production loudness. However, as some personalization may be done based on pre-selected preferences, the user may not be aware that the sound is personalized.

Harm and Offence Guidelines

Broadcasters follow a code of conduct and guidelines on a variety of content aspects that may cause harm or offence (See BBC editorial guidelines on harm and offence, [BBC, n.d.).

These include:

Managing Audience Expectations, Content Information, Labelling On-Demand and Digital Content, Scheduling for Television and Radio, Live Output, Language, Violence, Intimidation and Humiliation, Nudity, Sex, Abusive or Derogatory Treatment, Portrayal or groups or communities
Alcohol, Smoking, Vaping, Solvent Abuse and Illegal Drugs, Suicide, Attempted Suicide, Self-Harm and Eating Disorders, Imitative Behaviour, Tragic Events, Religion, Hypnotism, Exorcism, the Occult and the Paranormal, Flashing Images, Strobing and Images of Very Brief Duration, Acquired Programmes.

Branching narratives, alternative commentaries, and the re-ordering of content clips and chunks can cause challenges in both the creation and the monitoring of object-based media. Each object or branch creates a new version.

On the other hand, OBM has the ability to greatly ease the challenges experienced today with linear non-object-based media. With a good use of metadata elements can be categorised and only shown to appropriate audiences, with appropriate warnings.

For example, the creation of pre-watershead vs post-watershead versions could be automated and tracks that bleep out inappropriate language could be carried as an alternative to the original track, or even just the bleeps could be carried as a replacement object.

Sports fans often complain that the TV stadium audio lacks the voices of individual fans, which is a powerful aspect of the in-stadium experience, this is done intentionally to avoid unexpected profanity and other offensive language being broadcast. With appropriate use of objects and user opt-in these unfiltered voices could be carried as objects that are mixed in on request.

Some fundamental guidelines and models do need to be developed on how and what to monitor and regulate. This could range from ensuring that only the “default” experience follows all guidelines to some guidelines and checks being needed on all permutations of the content.

Data Privacy and Data protection

Content preferences, accessibility settings, choices in a branching narrative and the kind of audio/video equipment we have in our homes are pieces of data that are created and used when we consume object-based media.

It is also very interesting data that can say much about our lives and lifestyles and is potentially valuable to advertisers and retailers.

This raises questions of how to treat this data, who is granted access to it and how is it processed/sold/used.

If the content is streamed rather than broadcast, information on settings and personalization choices need to be provided to the operator to allow the personalization to be done, but how that data personal could, and might already be used today, is not clear.

If the personalization or object selection is done via a TV app or via the television itself, is the data the responsibility of the TV vendor, the TV app vendor, or the content provider?

As many apps and TV platforms already support content selection and recommendations and carry targeted ads, it is likely that the opt-ins and data protection regulation is already in place. It would however be prudent to ensure that guidelines are in place well in advance.

Accessibility, usability and equality of access

Object-based media brings many accessibility and access benefits ranging from AD, subtitle, and signing, through to alternative language tracks and dialogue enhancement.

In general, this is very positive and can help to provide more consistency, more choice and less fragmentation in the experiences being delivered.

There are however complexities that come with this choice and plethora of settings. Presenting the user with a complex UI which requires multiple button selections and parameters to be set is likely to result in a huge usability issue and furthermore raise additional accessibility issues. Accessibility experts are keen to point out that if a feature designed for users with sight loss needs screen menus and selections to be made, it is not usable by the people it is designed to help.

The use of default or automatic settings (e.g. setting AD once to “always on”) may help, as might new forms of user interactivity such as voice control. It is, however, crucial that the world of Objects provides equal access to the new features.

Advertising

Advertising has been one of the first markets to embrace object-based techniques with targeted ‘objects’ in the form of targeted ads already being deployed on most UK platforms including terrestrial TV.

Targeted ads can be seen as a close relative to branching narratives where a different chunk of video is played to different viewers based on a selection criterion. The main difference being that the content provider is in charge of the selection rather than the user.

This presents many opportunities to further blur the line between content and promotional material, where the targeted ads mechanism is used to deliver alternative endings and where ads and ad breaks are entwined into the narrative, making them less disruptive to viewing and more contextually sensitive.

By using the layering of objects, it is also possible to incorporate ads into the content itself. Companies such as Mirriad are utilizing AI to insert video objects containing ads directly into the programmatic content itself (Giatrakis), making the ads part of the programme, and not a time sliced event. Mirriad already lists Sky, Virgin Media and Channel4 as partners.

Similar experiments have been shown for audio where audio ads are inserted in a similar way to AD, for example ads during gaps in a sports match's commentary.

This of course has its challenges as well as its advantages. Integrated ads are less obtrusive but this means it's equally less clear what is or is not an ad.

Healthy Competition

There is no doubt that adding objects to your content is easier when you have control of your own distribution platform, and easier still if that distribution platform is capable of high bandwidth unicast broadband delivery and has a reliable and fast return path.

It is therefore not surprising that the most innovative use of objects is happening in the realm of pay TV operators and subscription OTT.

The BBC however also has its own strong self-controlled platform in the iPlayer which has therefore provided it with the same opportunities. This does however beg the question as to how smaller content providers and broadcasters can compete.

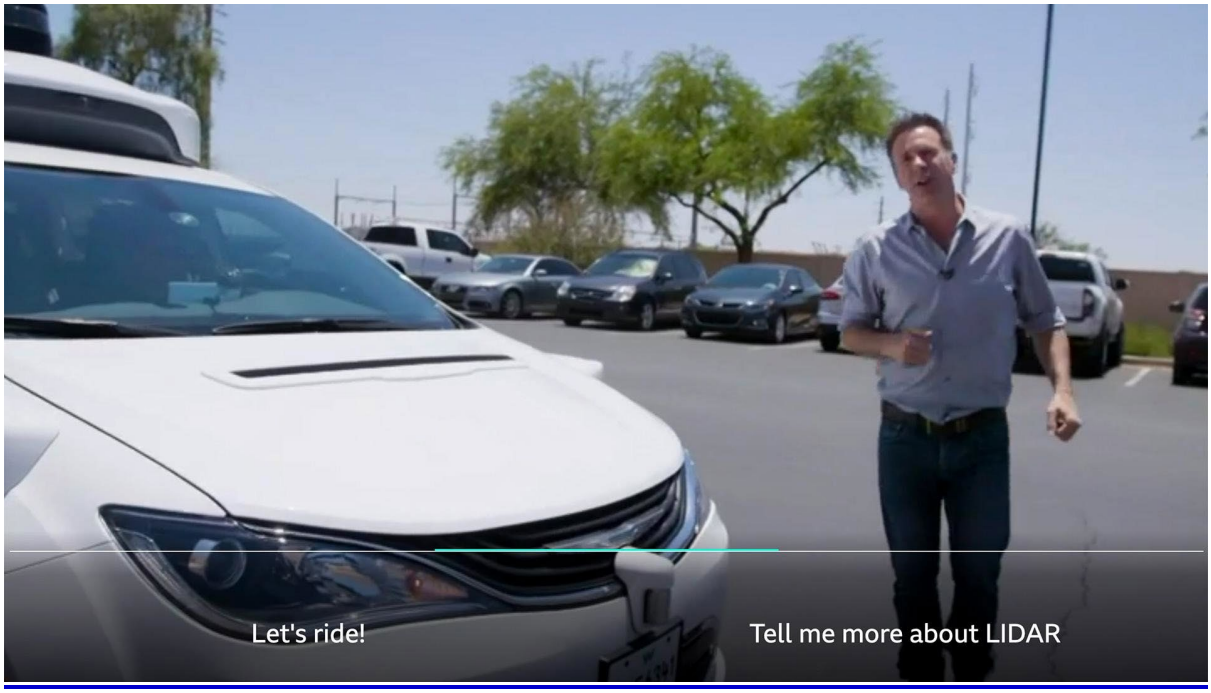
Developments thus far by the larger operators have been largely home-grown with tools developed in-house and experiments often not expected to yield a return on investment.

Standardisation is still in progress and commercial tools are often expensive or not available to the smaller players. This points to a growing gap in user experience between the smaller content creators and channels and the larger platform providers.

Open source tools might help and we are already seeing innovative partnerships forming between internet start-ups and innovative content providers. It is however worth considering if there are initiatives that can ensure that object-based media does not become another factor that exacerbates the gap between big and small providers and prevents market access to emerging providers.

ANNEX A: Additional media

Click 1000



A personalised, interactive episode based on your interests and tailored to how you want to experience the show.

The Mermaid's Tears
The World's first end-to-end Object-Based Media radio drama.

Doctor Who – Knock, Knock

A binaural version of an episode from series 10 made available on BBC iPlayer.

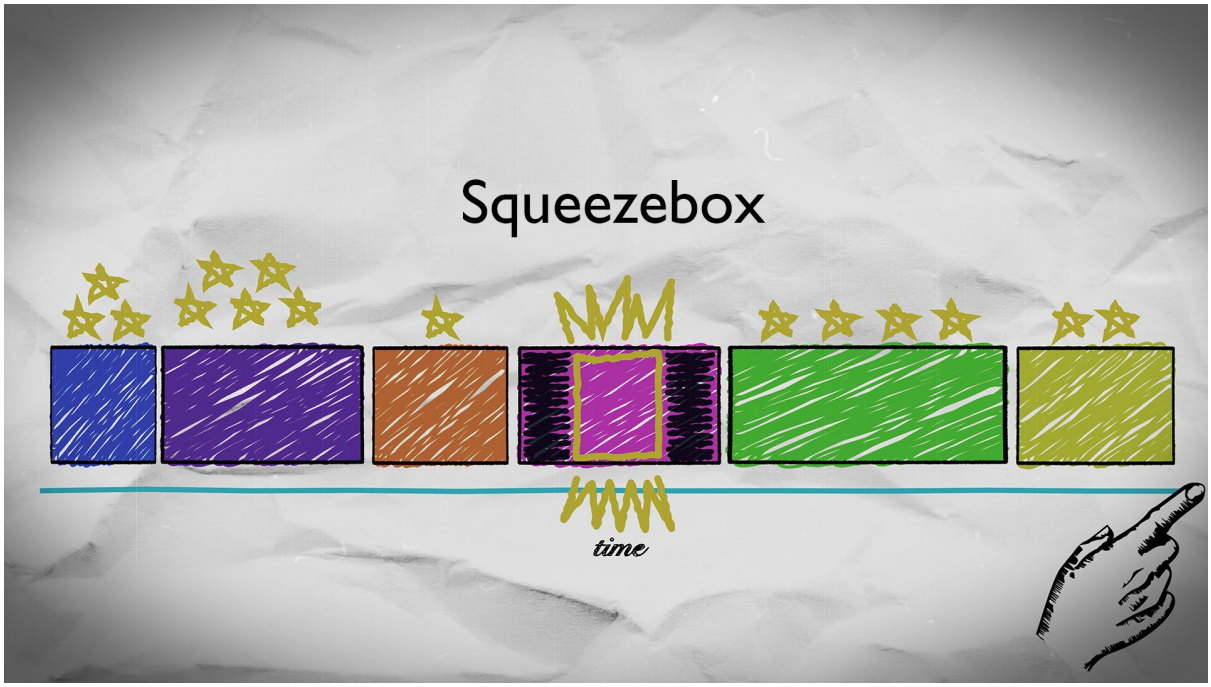
[Responsive Radio]
A radio documentary where audiences could listen to the complete story at a length that fit in with the time they had to listen.

[Forecaster]

An experimental Object-Based weather forecast.

[Story Explorer]
A new way to enjoy and understand drama and stories on the web.

Squeezebox



A production tool that can automatically create montages of news stories to any duration.

Newsbeat Explains/Atomised News

Piecing together the news for young audiences.

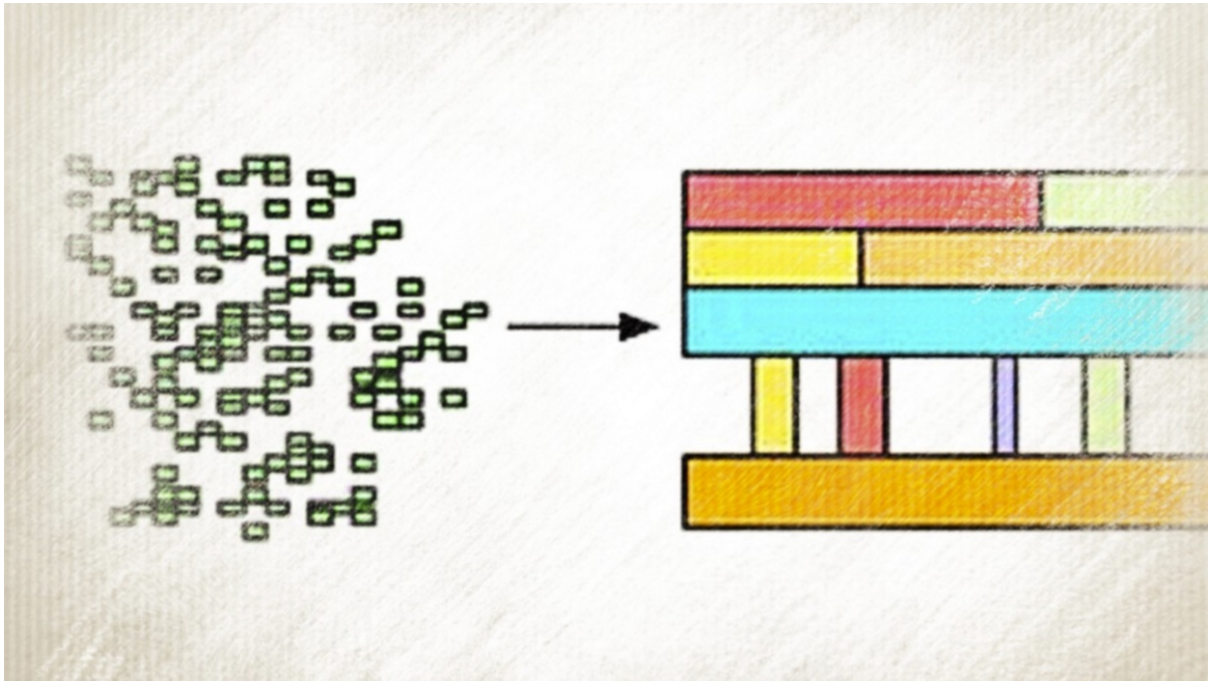
The Turning Forest

A magical binaural sound-led VR fairy-tale.

Visual Perceptive Media

Personalised video which responds to your personality and preferences.

Object-Based Composition – UMCP



R&D is exploring ways of making Object-Based Media to scale, infinitely repeatable and standardised.

Render Engine Broadcasting (REB)

Exploring ways to distribute and consume Object-Based Media experiences at scale and investigating how we broadcast live personalised AR experiences to millions of people.

ANNEX B: Bibliography

A Heist with Markiplier (2019) Directed by Markiplier.
ADM for absolute beginners (2020) Directed by D. Marston.

Amazon (2018) *4 ways to use X-Ray in Prime Video*. Amazon.com. [Online] [Accessed on 18th May 2021] https://www.amazon.com/primeinsider/video/pv-xray-tips.html.

Armstrong, M., Bowman, S., Brooks, M., Brown, A., Carter, J., Jones, A., Leonard, M. and Preece, T. (2017) 'Taking Object-Based Media from the Research Environment into Mainstream Production'.

Ballhaus, W. and Chow, W. (2020) 'Perspectives from the Global Entertainment & Media Outlook 2020–2024: Pulling the future forward: The entertainment and media industry reconfigures amid recovery'.

BBC (n.d.) *Section 5: Harm and Offence - Guidelines*. [Online] [Accessed on 19th May 2021] bbc.com/editorialguidelines/guidelines/harm-and-offence/guidelines/.

BBC R&D (2015) *Responsive Radio*. BBC Taster. [Online] [Accessed on 18th May 2021] https://www.bbc.co.uk/taster/pilots/responsive-radio.

BBC R&D (2019) *BBC MakerBox*. [Online] [Accessed on 18th May 2021] https://www.bbc.co.uk/makerbox/tools/storyformer.

BBC R&D (n.d.) *Cook-Along Kitchen Experience*. BBC R&D. [Online] [Accessed on 18th May 2021] https://www.bbc.co.uk/rd/projects/cake.

Beckett, S. (n.d.) *Celebrating 1000 Episodes with BBC Click - Case Study - BBC MakerBox*. [Online] [Accessed on 18th May 2021] https://www.bbc.co.uk/makerbox/case-studies/celebrating-1000-episodes-with-bbc-click-.

Bennett, J. and Murphy, A. (2020) *StoryFutures Academy Launches Skills for Immersive Experience Creation Report - StoryFutures*. StoryFutures Academy Skills for Immersive Experience Creation Report Launches - StoryFutures. [Online] [Accessed on 18th May 2021] https://www.storyfutures.com/news/skills-report.

BT (2020) *The enhanced BT Sport player*. BT.com. [Online] [Accessed on 18th May 2021] https://www.bt.com/help/bt-sport/the-enhanced-bt-sport-player.

Channel 4/YouTube (2021) *Channel 4 - YouTube*. [Online] [Accessed on 19th May 2021] https://www.youtube.com/user/channel4.

Churnside, A., Melchior, M. E. F., Armstrong, M., Shotton, M. and Brooks, M. (2014) 'Object-based broadcasting - curation, responsiveness and user experience.' *In International Broadcasting Convention (IBC) 2014 Conference*. Amsterdam, Netherlands: Institution of Engineering and Technology, p. 12.2-12.2.

DSR Corporation (2021) *IMF Conversion Tool*. DSR Corporation.

DTG (2020) *Next Generation Audio*. DTG. [Online] [Accessed on 18th May 2021] https://dtg.org.uk/resource/next-generation-audio/.

Epic Games (2021) *Film & Television*. Unreal Engine. [Online] [Accessed on 18th May 2021] https://www.unrealengine.com/en-US/industry/film-television.

Eurovision Song Contest 2019 - NGA demonstration - MPEG-H Audio (2019) Directed by EBU Tech.

Fetner, C. and Kensworthy, B. (2016) *IMF: An Open Standard with Open Tools | by Netflix Technology Blog | Netflix TechBlog*. [Online] [Accessed on 18th May 2021] https://netflixtechblog.com/imf-an-open-standard-with-open-tools-d85a665162a6.

Forever Holdings Limite d (2021) *In the room with Nile Rodgers*. [Online] [Accessed on 19th May 2021] https://hereintheroom.com/.

Google (2021) *how to fix a dripping tap - Google Search*. [Online] [Accessed on 19th May 2021] https://www.google.com/search?client=firefox-b-d&q=how+to+fix+a+dripping+tap.

Grylls, B. (2020) *You vs. Wild | Netflix Official Site*. [Online] [Accessed on 18th May 2021] https://www.netflix.com/gb/title/80227574.

Henninger, M. (2016) *Dolby AC-4 at CES 2016*. AVS Forum. [Online] [Accessed on 18th May 2021] https://www.avsforum.com/threads/dolby-ac-4-at-ces-2016.2287298/.

IBM (2018) *FOX Sports case study*. [Online] [Accessed on 18th May 2021] https://www.ibm.com/case-studies/fox-sports-aspera.

Introduction to IMF (2018) Directed by Netflix/Youtube.

IOC (2019) *Intel technology set to deliver several innovations during Tokyo 2020 - Olympic News*. International Olympic Committee. [Online] [Accessed on 19th May 2021] https://olympics.com/ioc/news/intel-technology-set-to-deliver-several-innovations-during-tokyo-2020.

Lomas, M. and Lumb, S. (2021) *Delivering Object-Based Media at Scale with Render Engine Broadcasting*. BBC R&D. [Online] [Accessed on 19th May 2021] https://www.bbc.co.uk/rd/projects/render-engine-broadcasting.

López, M., Kearney, G. and Hofstädter, K. (2020). Seeing films through Sound: Sound design, spatial audio, and accessibility for visually impaired audiences. *British Journal of Visual Impairment*, 1-28, DOI:10.1177/0264619620935935.

MBA@Syracuse (2021) 'Esports vs Sports | Can Esports Compete with Traditional Sports?' SYR-UMT. 16th February. [Online] [Accessed on 18th May 2021] https://onlinegrad.syracuse.edu/blog/esports-to-with-traditional-sports/.

Mirriad (2021) *Home*. Mirriad. [Online] [Accessed on 19th May 2021] https://www.mirriad.com/.

Mirriad end to end tech process (2019) Directed by M. Giatrakis.

NGA and the EAR production suite (2019) Directed by M. Weitnauer, M. Firth.

Object Based Broadcasting (n.d.). [Online] [Accessed on 18th May 2021] https://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP390.pdf.

Ofcom (2021) *Technology Futures: Spotlight on the technologies shaping communications for the future*.

Pearl - Enhancing Audio Description - Full Film with Binaural Soundtrack (2020) Directed by Palumbo and Feng. (EAD soundtrack by M. López and G. Kearney).

de Pomerai, R. (2019) *Will IMF replace AMWA AS-11? | DPP - the media industry's business network*. [Online] [Accessed on 18th May 2021] https://www.thedpp.com/news/will-imf-replace-amwa-as-11.

PWC (2019) 'Getting personal: Putting the "me" in entertainment and media' p. 27.

RNIB (2020) *Accessible televisions*. RNIB - See differently. [Online] [Accessed on 19th May 2021] https://www.rnib.org.uk/sight-loss-advice/technology-and-useful-products/technology-resource-hub-latest-facts-tips-and-guides/accessible-televisions.

Ruppel, W. (2020) *IMFTool*.

SBTVD (2021) 'Fórum SBTVD | TV 3.0 Project.' [Online] [Accessed on 19th May 2021] https://forumsbtvd.org.br/tv3_0/.

Sky (2021) ‘Smokefree | AdSmart from Sky Success story.’ AdSmart from Sky. [Online] [Accessed on 19th May 2021] https://www.adsmartfromsky.co.uk/case-studies/smoke-free-and-sky-adsmart/.

Sky Studios, Vertigo Films, and Charisma A.I. (2020) *Bulletproof: The Interrogation*. [Online] [Accessed on 19th May 2021] https://bulletproof.charisma.ai/.

Stornaway site story map clip (2021) Directed by Stornaway.io.

Tellyo (2021) *Cloud-based video editing, production, streaming, distribution | Tellyo*. [Online] [Accessed on 19th May 2021] https://tellyo.com/.

Ursu, M., Smith, D., Hook, J., Concannon, S. and Gray, J. (2020) ‘Authoring Interactive Fictional Stories in Object-Based Media (OBM).’ *In ACM International Conference on Interactive Media Experiences*. Cornella, Barcelona Spain: ACM, pp. 127–137.

Vailshery, L. (2021) *Number of voice assistants in use worldwide 2019-2024 | Statista*. [Online] [Accessed on 19th May 2021] https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/.

Whitehead, F. (2018) *Black Mirror: Bandersnatch | Netflix Official Site*.

ANNEX C: Acknowledgements

This report was commissioned in early 2021 by Simon Parnall, Principal Advisor, Broadcast Technology and Sanya Osisanya, Senior Technology Advisor at Ofcom's Technology Group.

The report was authored by Elfed Howells, CEO of Macroblock Ltd. and Dr David Jackson, Lecturer and Research Associate at the School of Digital Arts (SODA) within Manchester Metropolitan University.

The authors would like to offer thanks to the following for their generous contributions of time and knowledge to inform the findings of this report:

Adrian Murtaza, Fraunhofer IIS

Adrian Woolard, BBC

Andy Bell, Channel 4

Ben Shirley, Salsa Sound

Caroline White, University of York

Chris Johns, Sky

Chris Pike, BBC

Chris Walker, Bright White Ltd.

David Holiday, DTG

Doug Williams, BT

Graham Mills, Consultant

Guy Gadney, Charisma.AI

Ian Kegel, BT

Jesal Vishniram, RNID

Johnny Johnson, National Film and Television School

Jonas Roden, Dolby

Judy Parnall, BBC and EBU

Kirsty Fairclough, Manchester Metropolitan University

Mariana Lopez, University of York

Matthew Brooks, BBC

Orpheus Warr, Channel4

Renne Rummel-Mergeryan, Netflix

Rhun Roberts, BBC Cardiff

Ru Howe, Stornaway

Rupert Brun , Fraunhofer IIS

Simon Fell, DTG

Simon Tuff, BBC

Sonali Rai, RNIB

Toby Heys, Manchester Metropolitan University

Tom Griffith, ITV

Yvonne Thomas, DTG