Audiovisual Media Formats for Browsers Community Group

Meeting: 14 September 2023

Agenda

- Welcome and introductions (10 min)
- Mission and charter of the group (15min)
- Proposal for Addition to Media Capabilities Specification - Timo Kunkel, Dolby (30min)
- Use cases for NGA personalization - Wolfgang Schildbach, Dolby (15-30min)
- Future path and closing remarks (5 min)

Attendees

- Please sign in here!
- Wolfgang Schildbach (Dolby Laboratories)
- Chris Needham (BBC)
- Nigel Megitt
- François Daoust
- Kaz Ashimura (W3C)
- John Riviello (Comcast)
- Pat Griffis
- Eric Carlson
- Jean-Yves Avenard
- Christoph Guttandin
- Hongchan Choi (Google Chrome)
- Bernd Czelhan
- Rachel Yager
- Paul Adenot
- Rijubrata Bhaumik (Intel)
- David Singer (Apple)
- Xiaohan Wang (Google Chrome)
- Jiantong Zhou (Huawei)

Minutes

- Introduction and Welcome
- Slides: https://github.com/w3c/avmedia-formats-cg/blob/main/meetings/2023-09-14-slides.pdf
    - Started ~1yr ago
    - Agenda
        - Have two proposals to discuss today

                - Addition to MCA
                - NGA personalization usecases
            - Future usecases
            - Next steps
- Mission and Charter
    - Group started one year ago, gathered feedback since then
    - Feedback is that this group can discuss both commercial and open formats
    - Full charter is on github
    - Want to provide a forum to discuss relevant open and commercial formats for all devices, smart TVs
    - How can the formats be called from web environment
    - How to reference the formats
    - Look at adjacent fields (film industry, broadcast, OTT) and get their input too
    - Make sure that these fields are covered more than in other groups
    - Want to discuss challenges to current Web APIs
    - Want to report on relevant trends and developments, will be raised in relevant issue trackers
    - We are not intending to create any specifications, not under current charter
    - No plan to create test suites or such
    - Based on current CG setup, cannot create normative specs but can create community reports
        - Additions to APIs
        - New APIs
        - Look at Tech landscape
        - Things to look out for, bring to other group's attention
    - We need to collaborate, ref: Liaison
        - Media Entertainment IG & CG
        - Audio WG
        - Color Web CG
        - Immersive Web CG
        - External groups
            - Please let us know who else to reach out for for external liaisons
    - Charter has been deliberately setup to be similar to the MEIG
        - But we are not writing any specs
        - Have discussed how to evolve, we may fold the two groups into one such that there is one group dealing with media related use cases
        - As it is, they are distinct but closely related
    - Please bring topics that fit the scope
        - Specifically if the topics can't easily be brought into other groups
    - We have two proposals at this time
    - If we get agreement that the proposals are a good idea to drive forward, we will produce a report to input into other groups
    - (Eric Carlson) Q: structure is set up such that companies don't have to join a WG and agree to terms?
    - (Timo Kunkel) No: it is an entry path to have a voice without being members. Not meant to circumvent anything, but rather to tap into new potentials. Be able to get input from media industry.

    - (Pat Griffis) This forum people can come and discuss how to make the web interoperable for open & commercial formats. Output would go to other groups, depending on the topic, and under their terms the ideas would be put into documents.
- Proposal 1 (Timo Kunkel, Dolby Laboratories)
    - Slides: https://github.com/w3c/avmedia-formats-cg/blob/main/meetings/2023-09-14-media-capabilities.pdf
    - Addition to the Media Capability Specification
    - (insert link to presentation here)
    - Overview of MC API:
        - Provide API to enable web apps pick optimal format given the device capabilities
        - There are capabilities related to HDR
            - Colour gamut, transfer function and HDR metadata type are in
            - Various SMPTE specs are referenced
            - First two are sufficient but the metadata format is not
                - Some open formats are available, but no commercial formats
                - Dolby Vision is not included even though widely available, in mobile devices, computer displays, and TVs
                - Also adopted by many content providers
                - For this, Dolby thinks it is a desirable property
                - Perception is that identification of Dolby Vision is possible already, btu closer inspections shows it is not
                    - SMPTE 2094-10 datatype is a partial subset of Dolby Vision, and without explicitly signaling, apps cannot identify capabilities for the full stream.
                    - Cannot be identified as codec either, because Dolby Vision can be used with many different codecs
                - 2020: Initial discussions, but put on hold not to delay spec release
                - 2022: more discussions (reference github in the presentation)
                - 2023: More progress in both open and commercial formats. Identification means for Dolby Vision is now needed.
                - Proposal is “dvmd” to identify Dolby Vision streams
                    - Only establishes the plumbing
                    - To identify whether the rendered is present in the playback stream
                - Documentation of the datatype
                    - Documents will be made available on a platform, end of Sep 2023
                    - Sample code & streams

- Question to group
  - Does the proposal have merits?
  - Are there concerns?
- (Pat) HDR has many flavours these days.
  - HLG
  - PQ
  - But also different codec flavours.
    - HEVC most important,
    - AVC also supported
    - Others are on the horizon
    - HEVC + PQ combination
    - HDR support in W3C is important and not unique to Dolby
    - It is becoming mainstream
    - We want it to be more interoperable
    - There may be other topics.
    - Note that 70% of evening content is media
- (Jean-Yves) This may open the door to thousands more formats
  - Could we consider a format that makes the type parametrizable like “custom=XXX”?
- (Timo) There are not that many relevant codecs
- (JY) the custom identifier itself does not have to be mentioned in spec
- (Francois) One approach is to have a registry, so that user agents understand. It is weird to have an identifier that preferences one specific vendor. Without a specification, if that vendor goes away, nothing is left. SMPTE is different in that third parties can implement.
- So the custom approach would be reasonable, the codec could be hooked in.
- (Timo) Traditional thinking is that reference is standardised. But this format does not go away, it is widely deployed. Is there a way to reference things of commercial nature in the W3C specs? Is this a no-go or is there a way.
- (Francois) can we turn the enum into something that uses a registry? Are there implementations already?
- (Jean-Yves) this sounds reasonable
- (Kaz, W3C) Would W3C manage a registry for HDR related methods, or can we expect some other organisation who’re already working on something similar?
- (Eric) Timed Text WG has a similar plan.
- (Nigel) there is a plan for a registry, if WG agrees on rules then it can be managed, requirements for references.

- (Chris) Went through same thing in Web Codecs. Could be done.
- (JYA) similar to MSE spec
- (Rishubrata) This is a common path for commercial/proprietary formats, would be in support
- (JYA) would help for existing formats even because could contain more explanation
- (Chris) Two ways to do registry: could be separate document or embedded in the spec.
- (Nigel) There is a different review process
- (Francois) Registry is not normative as such, it is not subject to patent policy. Just a list of key/value pairs. Does not create requirements per se
    - https://www.w3.org/2023/Process-20230612/#registries W3C Process - 6.5 The Registry Track
- (Jiantong) Should we consider more commercial formats. In China, have HDR10+, Dolby Vision and HDR Vivid. Also, PQ; and HLG are supported. Lots of different formats. Maybe wider support of market used formats can be included.
- (Nigel) Q: Which HDR formats need additional metadata?
- (Jiantong): Dolby Vision for sure, and HDR VIVID which is used in China also needs.
- (Timo) Others have been proposed as well but are not widely supported. A registry might serve the purpose, could be an option. At the moment would only contain two entries?
- (Francois) Plus the three that are there already.
- (Nigel) Of the formats that require metadata, which ones do we expect to be played back on user agents
- (Xiaohan) Dolby Vision needs special hardware. The browser would have to do some feature detection to decide if playback is supported
- (timo) Level of commercial support is relevant. We should find mechanisms to take that into account. This (Dolby Vision) is just the first format to come along.
- (Pat) Sounds like there is agreement to have this. Registry makes sense because it scales. This will be a growing reality. If we don't do it here, there is risk that browsers diverge. Codec support is up to the device and outside of W3C but we want to create interop. If the group agrees, we could propose to the MEWG to producing a registry.
- (Chris) criteria for inclusion would have to be decided there. Lower-level detail of what the actual value is… is to be decided.

- (Kaz) However, still wondering if there are other registries already, created by other orgs? We should survey these other orgs, and should work with them if any.
- (Nigel): Yes, is there something in mp4ra? Maybe it should sit there not here?
- (Timo) Not at the moment, we may have missed something though. There are some code points but don't serve the purpose
- (Pat) this would give the option to add other formats, like HDR VIVID format, HDR10+ . Expectation would be commercial availability, scale, and longevity.
- (Nigel) why does browser even know this?
- (Francois) because this is about query.
- (Eric) web app needs to query prior to playback.
- (Nigel) why does the browser need it?
- (Chris) Does what the page does depend on the values that are within that metadata? Does knowing "there's DV metadata available" give the page enough info?
- (did not capture part of the discussion)
- (Timo) Dolby Vision has many different modes, this cannot be captured with the current parameters.
- (Wolfgang) With DASH streaming, two adaptation sets, one with HEVC and DV metadata, and another format, the page needs to decide which adaptation set to page. They only select the backwards compatible value if it's known to be playable.
- (Nigel) I can see that if the page needs to decide what it can play then it might need to know what can be processed, but is this about signalling media content or device capability? If it's DASH signalling of media content in Adaptation Sets then that seems like a DASH issue.
- (Xiaohan) That is how it works, if two datatypes are contained in MPD. Are the mimetype and metadata type orthogonal?
- Eric: It's possible to decode some streams without making use of all the data
- Xiaohan: choosing happens at decoding time, how does the browser choose which one to do? In MSE when you append the buffer, you give the mimetype, as a signal.
- Wolfgang: We assume they'd go through the same codec on the system
- Xiaohan: Dolby profile 8 can be done as pure HEVC. Alternative proposal could be to use the mimetype in MSE. How can the player tell the browser which mode to use, which is after the MC API query

        - Eric: Assumption is the UA will decode at the highest fidelity the system is capable of. This signalling is for the page to decide which is most efficient for the system
        - Timo: Action is how to setup a registry.
        - As a CG, would like to have more regular meetings
        - What is a good cadence?
        - Take aways
            - Look at the options
            - Dolby will consult their own experts
            - Sounds like this is a good avenue
        - Nigel: Explaining the sequence of what thing/actor needs to make what decisions when, and what information is available to support those decisions, will be good. To determine whether we are making the right design.
        - Timo: This will be in the report.
        - Pat: Likes the registry idea. Will future-proof the spec. When we get into the audio formats, could bump up against the same questions.
        - Chris: Next step is to develop the doc. Describe how this operates, what gets signaled, what gets passed into MCA and into MSE.
- Next Generation Audio use cases
    - Wolfgang: Reporting for a group of companies, BBC, Dolby, Fraunhofer IIS have discussed use cases for personalisation that could be improved, needing W3C support. I'll run through the use cases to explain and report on our progress
    - Many use cases relate to a11y, and some about more flexible experiences. Dialog enhancement improves intelligibility, it's a personalisation option, on/off. The degree of how much you want to improve personalisation is a user choice.
    - Next is selection of audio elements, there's a choice that can be made client-side, e.g., between languages or, more complex, the components going into a mix. AD mixes main audio with a description of what's in the picture. Another case is narrative importance, the sounds that tell a story, e..g, an alarm or a doorbell. These could be lots of different sound elements, some vital to a scene. Can create challenges for intelligibility, so NI focuses on those central to a story.
    - Gain interactivity refers to being able to individually control gains on audio elements, giving prominence to certain elements. Controlling gain could include -Infinity, so switching it off
    - Position interactivity, relative to screen or listener. This can be made a user choice, improve spatial separation. Football match with different audiences in different parts of the scene
    - Nigel: *raises eyebrows at idea that the audience sounds could be moved so they don't match the video image*

- Bernd: other use cases, e.g., an AD moving to the person who needs the AD to help the understand the scene
- Wolfgang: Preselections is a basic form of personalisation. Preset static mixes of objects controlled by the content provider. Depending on the page, the user can deviate from that mix and the defaults
- Next step for the group would be for the group would be how this could look in an API.
- Chris: We don't have an API proposal for this but we've been working on the same use cases and are seeking feedback on interest, and if there is interest, how we might surface it.
- Eric: How would it work, as a user preference? On mobile, volume control is strictly controlled by the user. There's a volume attribute on a video element, but script isn't allow to change the volume unless it's in response to user gesture. So something that automatically changes the volume of elements in a page wouldn't work unless run in response to a user gesture. How do you imagine that would work. Where's the preference, what is the granularity of the preference? How to prevent it being gamed, e.g., ads pumping up the volume
- Wolfgang: I understand the concern. Preference isn't really a system preference, implying automatic behaviour. It is a personal preference and expressed through gesture. The page would have UI elements.
- Xiaohan: HTMLMediaElement has volume, playback rate already, this seems similar. Media element has audio tracks, so is this extending that to add per-track volume?
- Chris: That's an idea we've had, yes
- Bernd: I was thinking of a programmatic API to build your UX. For example it would say it has these attributes
- Nigel: So the media signals something about what's offered, the page can query that, show UX
- Chris: There's metadata to describe that and collaboration to define a common set between codecs
- Nigel: For accessibility, …
- Wolfgang: In simplest use cases, choose a preselection. DASH deals with it on a manifest level, there's an accessibility element.
- Nigel: That's about selecting a different representation, so there is one set of resources per representation.. A premixed audio track normal one, or a premixed track with AD. But this seems different with individual components exposed within a single audio resource, with the page or user able to adjust the audio rendering of the content within that resource, e.g. changing the levels
- Wolfgang: There is that, but preselections are kind of orthogonal. One stream, three preselections
- Eric: So it's a description of how to configure the playback the media assets in the manifest?
- Wolfgang: Yes, and how to configure the playback system to pick elements contained within one stream. Preselection elements can have varying complexity: play multiple adaptation sets side by side. It isn't done today,

means opening up different buffers in MSE. Or preselection with elements in one stream, so you configure the decoder and player.
    - Kaz: On intelligibility and accessibility, if this is about adaptation of teh speech signal, that's fine. How about conversion of the speech or regeneration itself,e.g., on a mobile phone. Or translation of speech, or changing the voice type. A "clear voice". Make sense to clarify the scope of the proposal.
    - Nigel: There is a tension there. I worked on the idea of passing both the text and audio mixing instructions for AD to the player. The page could choose whether to expose the text to assistive technology or also do mixing with audio. Would be helpful to have a model of which components in the web platform are doing the mixing, what ability is there to script. Middle ground, put them into a web audio context, have them labelled. They could have important a11y feature labels, so if there's a system setting that handles those it can be handled in a different way.
    - Pat: A11y is important to W3C, is that the calling card for this activity. Lots of features that make it interesting, but those are value add on top of a11y. I'm hearing general interest. Get the a11y part right, go from there
    - Eric: Devil is in the details
    - Chris: are we asking for a concrete proposal
    - Eric: yes, have concrete proposal to understand how this works
- Wrap up
    - Timo: We'll put minutes in the repo and decide on a next meeting time. Will send a Doodle poll in the coming weeks. In the meantime, we'll do our homework on action items
    - Pat: Meeting cadence?
    - Timo: Start with monthly, then see if we need to go faster or slower.
    -