

Slides: <https://www.slideshare.net/christian.timmerer>



From Pixels to Semantics: Where Streaming Trends Meet the Container

Christian Timmerer, CIO | Head of Research and Standardization at **Bitmovin**,
Univ.-Prof. at AAU, Directors at ITEC, Director at CD Lab ATHENA
Klagenfurt, Austria

July 14, 2026



Presenter



2003: MSc CS (Dipl.-Ing.)

2006: PhD CS (Dr.-techn.)

2012: Co-founded Bitmovin

2014: Habilitation (Priv.-Doz.) & Assoc. Prof.

2016: Dep. Director @ ITEC/AAU

2019: Director @ ATHENA

2022: Univ.-Prof. for Multimedia Systems

2024: Director @ ITEC/AAU

Christian Timmerer

Web: <http://timmerer.com/>

Univ.-Prof. at Alpen-Adria-Universität Klagenfurt

Director ITEC

Director CD Lab ATHENA

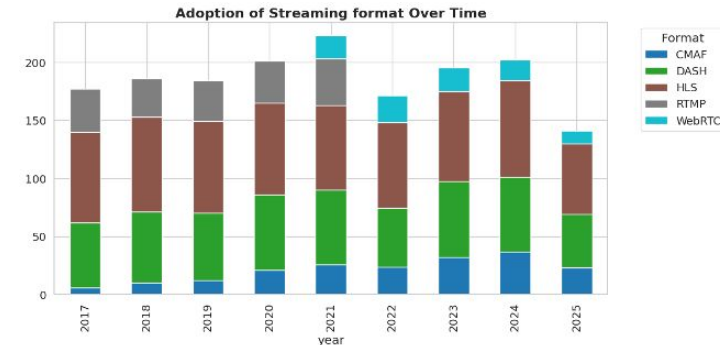
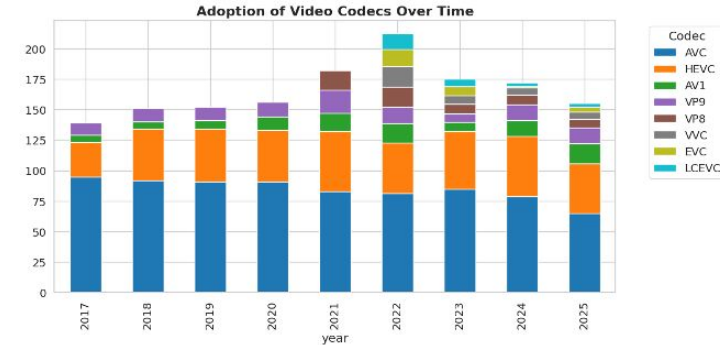
CIO | Head of Research and Standardization at Bitmovin

Agenda

- Introduction
- Efficiency & Sustainability
- Latency
- Artificial Intelligence
- Synthesis
- Conclusions

Introduction

- Video streaming accounting for >65% of Internet traffic
- Bitmovin Video Developer Report (9th ed. avail., 10th ed. coming soon) – see m74403 (MPEG152)
 - Controlling cost (bandwidth, storage) is the top challenge at 38%, narrowly overtaking ad insertion (35%)
 - Live pipelines focus on reliability and low latency, while VOD emphasizes scale, automation, and efficiency
 - AI usage for audio transcription/speech-to-text (46%), personalization (35%), tagging/categorization (35%), and recommendations (33%)



Sources:

2024 Global Internet Phenomena Report. Technical report, AppLogic NETWORKS.
 2025 Bitmovin Video Developer Report (9th ed.)

Efficiency – Bitrate Ladder Construction

Content-awareness (aka per-title)

- Take into account perceptual, just-noticeable differences (JND), and user satisfaction based on visual content complexities ((E)VCA)

Context-awareness

- Take into account network conditions and device information

Multiple parameter optimization space

- Bitrate, resolution, framerate, bit depth, color subsampling, .



Increase flexibility for bitrate ladder construction across various dimensions with limited support elsewhere



Enable quality-related metadata to travel alongside media data but comes with add'l costs

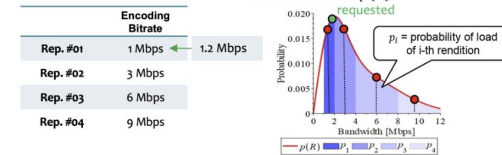
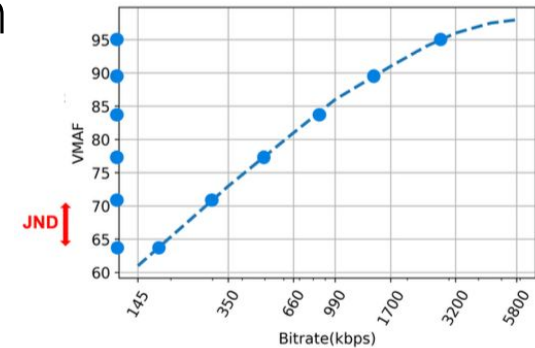


TABLE VI: Average results of the encoding schemes compared to the HLS CBR encoding.

Method	BDR_P	BDR_Q	BD-PSNR	BD-VMAF	ΔS	ΔT
Bruteforce (2 VMAF JND) [3], [13]	-23.09%	-43.23%	1.34 dB	10.61	-25.99%	4732.33%
Bruteforce (4 VMAF JND) [3], [13]	-28.15%	-42.75%	1.70 dB	10.08	-59.07%	4732.33%
Bruteforce (6 VMAF JND) [3], [13]	-25.36%	-40.73%	1.67 dB	9.19	-70.50%	4732.33%
OPTE CBR [11]	-17.28%	-22.79%	0.98 dB	3.79	0.07%	15.74%
PPTE (2 VMAF JND) [7]	-11.06%	-16.65%	0.87 dB	2.18	10.18%	105.73%
PPTE (4 VMAF JND) [7]	-10.44%	-15.13%	0.91 dB	2.39	-27.03%	10.19%
PPTE (6 VMAF JND) [7]	-12.94%	-17.94%	0.94 dB	2.32	-42.48%	-25.35%
HLS cVBR	-35.25%	-32.33%	2.09 dB	6.59	-9.39%	1.64%
OPTE cVBR	-34.42%	-42.67%	2.90 dB	9.51	-1.34%	62.73%
JTPS (2 VMAF JND)	-14.25%	-29.14%	1.36 dB	7.82	23.57%	184.62%
JTPS (4 VMAF JND)	-18.41%	-32.48%	1.41 dB	8.31	-56.38%	26.14%
JTPS (6 VMAF JND)	-18.80%	-32.59%	1.34 dB	8.34	-68.96%	-18.58%

Results: 10.1109/TCSVT.2023.3290725

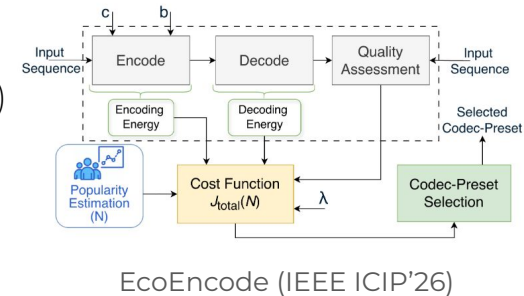
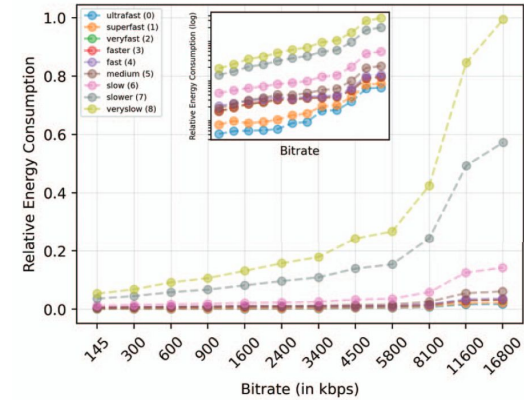
Sustainability – Bitrate Ladder Construction

Energy-awareness

- Take into account energy consumption of encoding and decoding to determine bitrate ladder across a multiple parameter optimization space
- Various tools & datasets available [\[ATHENA Reproducibility\]](#)
- EcoEncode: popularity-aware codec/preset selection: trades encoding energy against compression efficiency (up to -99% energy for 1-4 VMAF on rarely-watched content)



Enable energy metadata carried in manifests and containers, decoder-energy-aware adaptation  
CMCD/CMSD-style signaling  industry trend 



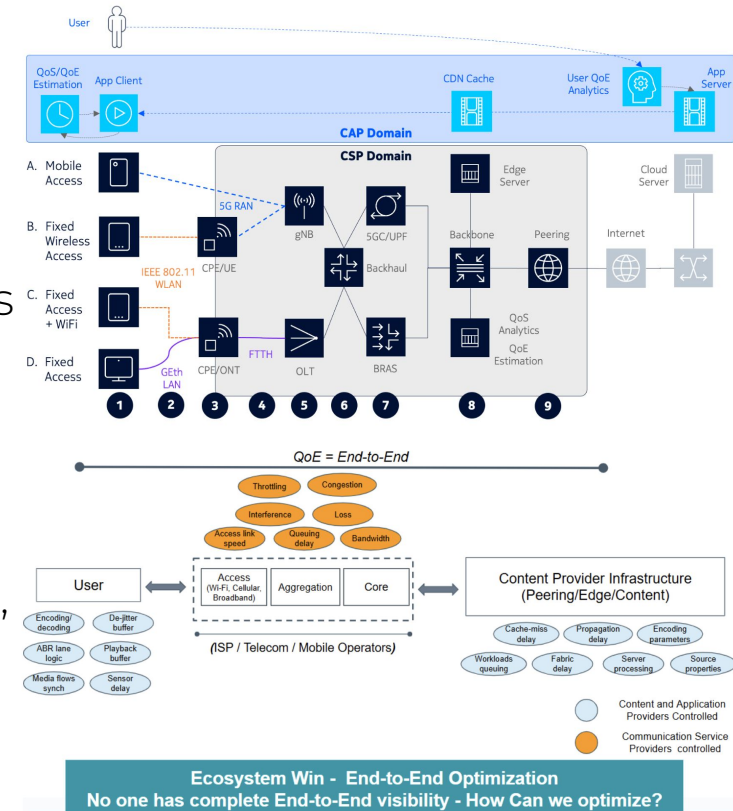
Bridging the Gap: CAPs <=> CSPs

The industry is converging on cross-layer cooperation

- QoS and QoE information at the interface between Content and Application Providers (CAPs) and Communication Service Providers (CSPs)



Systems takeaway: standardize the metric vocabulary (ITU-T / VQEG)  
bind it into the client/server channel (DASH, CMCD/CMSD) 



Latency

The low-latency boundary is moving

- The live streaming trilemma: latency vs scale vs simplicity — pick two!
- WebRTC: sub-second but poor fan-out
- DASH/HLS: CDN-scale but seconds of delay
- MoQ's pitch: collapse the trilemma; sub-second at CDN scale

**The new kid
on the block!**

Sub-second interactive

WebRTC · MOQ

Live betting
Auctions
Live commerce
Second-screen

Low-latency live

LL-DASH / LL-HLS

Sports
News
Live events

Linear + VOD

DASH / HLS

FAST channels
Catch-up
Box sets

Ultra-low latency

Latency-tolerant

Latency

O “LL-DASH / LL-HLS & MoQ”, Where Art Thou?

	LL-DASH / LL-HLS HTTP adaptive streaming	MoQ Media over QUIC (IETF)
Transport	HTTP / TCP	QUIC (UDP)
Delivery model	Request–response (pull)	Publish/subscribe + relays
Latency	~2–6 s practical	Sub-500 ms (target)
Scale	Proven, CDN-scale	CDN-scale via relays
Maturity	Universal, decade-proven	Draft, months from RFC
Best fit today	VOD, linear, most live	Sub-second interactive
Standards home	MPEG / DASH-IF / CTA	IETF (+ MPEG format)

Shared substrate: CMAF / ISO BMFF

one packaged asset, both delivery worlds

What it means for MPEG Systems?

- Format continuity is the prize: keep CMAF/ISO BMFF as the shared substrate
- Roadmap question isn't "who wins", it's interoperability and a common format so the two coexist



MPEG owns CMAF/ISO BMFF ✓
 IETF owns MoQT ✓
 CMAF-to-MoQT object mapping ✗ (incl. timing model, group boundaries, init-segment handling, and encryption signaling, ...)

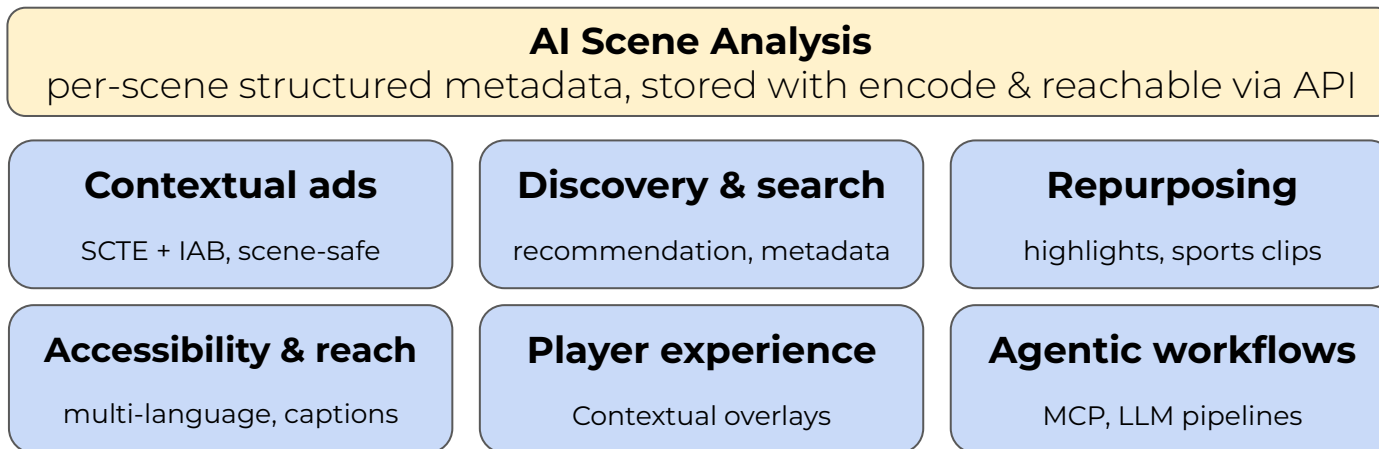
IETF 126 Vienna

18 Jul 2026 - 24 Jul 2026

Artificial Intelligence

AI moved inside the pipeline, not beside it

- AI is now in production across streaming, not experimental: e.g., accessibility/transcription, tagging, personalization, quality optimization
- Two distinct things AI does to media: *understand it* (turn it into structured metadata) and *synthesize it* (generate/reconstruct pixels)



Artificial Intelligence

The catch: AI metadata doesn't travel

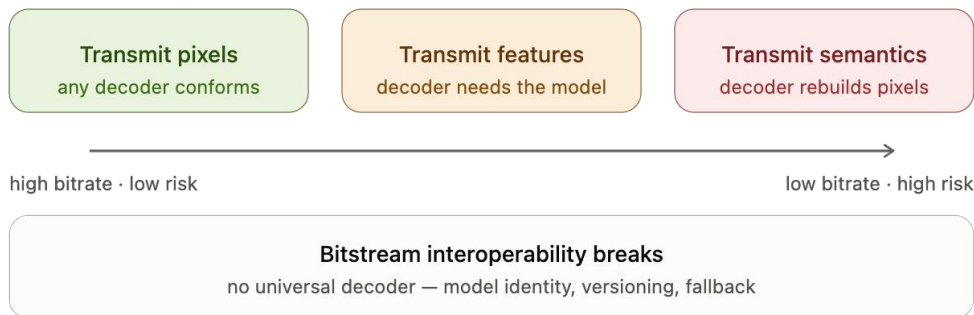
- The metadata is rich but proprietary: a vendor JSON schema, English-first with LLM translation, reachable through one vendor's API
- The ad layer is standardized (SCTE cues, IAB taxonomy), the semantic/narrative layer is not. There's no standard way to express AI scene metadata portably. Every downstream integration is bespoke.

➔ A standardized timed  semantic-metadata schema  ( MPEG-7 ) and carriage 

Artificial Intelligence

When AI stops describing and starts reconstructing

- As AI shifts from describing media to reconstructing it (neural and generative codecs, edge VLMs that ship semantics instead of pixels) bitstream interoperability itself breaks: the decoder needs the matching model



→ Model interoperability: neural-network coding plus signaling of which model/version  

The systems questions it forces open:

- Model identity and versioning: how does a stream signal which model reconstructs it?
- Model distribution: neural-network coding to ship and update decoders, not just bitstreams.
- Conformance, redefined: from "bit-exact output" to "acceptable perceptual/functional equivalence," which the whole test-and-certification apparatus is not built for.
- Graceful fallback: what a non-AI or wrong-model decoder does instead of failing.
- Provenance: once pixels are generated rather than captured, authenticity signaling (C2PA) stops being optional.

Synthesize

Three pressures, one layer

- Efficiency & sustainability metadata with binding into DASH, CMCD/CMSD (but also elsewhere)
- A normative CMAF-to-MoQT object mapping so one packaged asset serves both request/response and pub/sub delivery
- (a) A standardized timed semantic-metadata schema and carriage with taxonomy binding, so AI scene metadata travels across tools.
(b) AI-reconstructed media has no universal decoder; interoperability moves up to the systems layer.

Conclusions

The systems layer is the next inflection

- The next inflection isn't a codec or a protocol — it's whether the systems layer can carry what AI and the network already know — observability is a key!
- Efficiency, low-latency live, and AI-native media all converge on the same gap: the signal exists in production, the standardized carriage doesn't — let's get the party started!



Thank you for your attention