Media & Entertainment on the Web

(R)Evolutions of the Web platform

W3C WebEvolve Conference
28-29 May 2024 - Shanghai

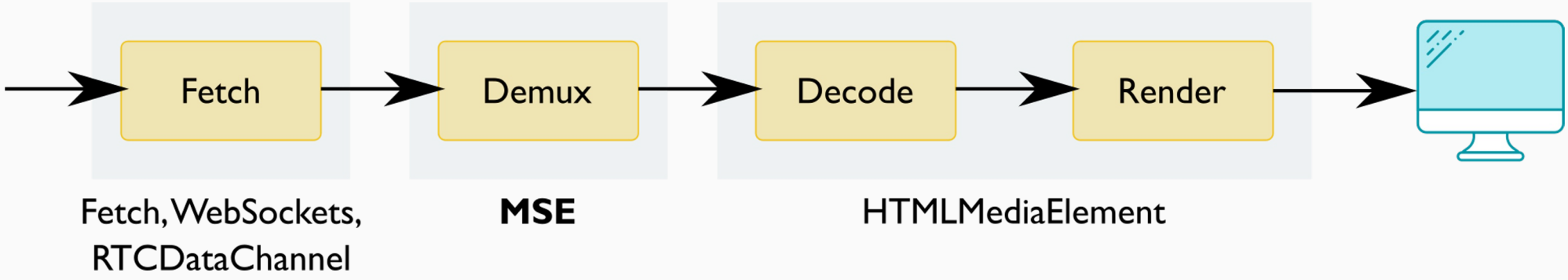
François Daoust - fd@w3.org - @tidoust

Goal: Discuss ongoing/upcoming/possible Web (R)Evolutions for Media & Entertainment

- Real-time media on the Web
- Web Games
- Web & AI
- The metaverse


Real-time media
on the Web
8-19 November 2021
Opening Remarks
https://www.w3.org/2021/03/media-production-workshop/

Web media distribution ~~today~~ yesterday



- Decode & Rendering: `<video>` in HTML
- Demux / Adaptive streaming: Media Source Extensions (MSE)
- Content protection: Encrypted Media Extensions (EME)
- Captions: TTML or WebVTT or…

Media on the web: Ongoing evolutions

- Address device fragmentation
  - `ManagedMediaSource` in MSE
  - Media Capabilities
  - Picture-in-Picture
  - Web Media APIs & tests (collaboration with CTA WAVE)
- System integration
  - Media Session
- HDR/WCG support on the web
  - HDR support in canvas
- Better support for ads customization
  - Codec Switching feature in Media Source Extensions (MSE)
- Open second screen support
  - Open Screen Protocol

Real-time and interactive experiences

Ongoing

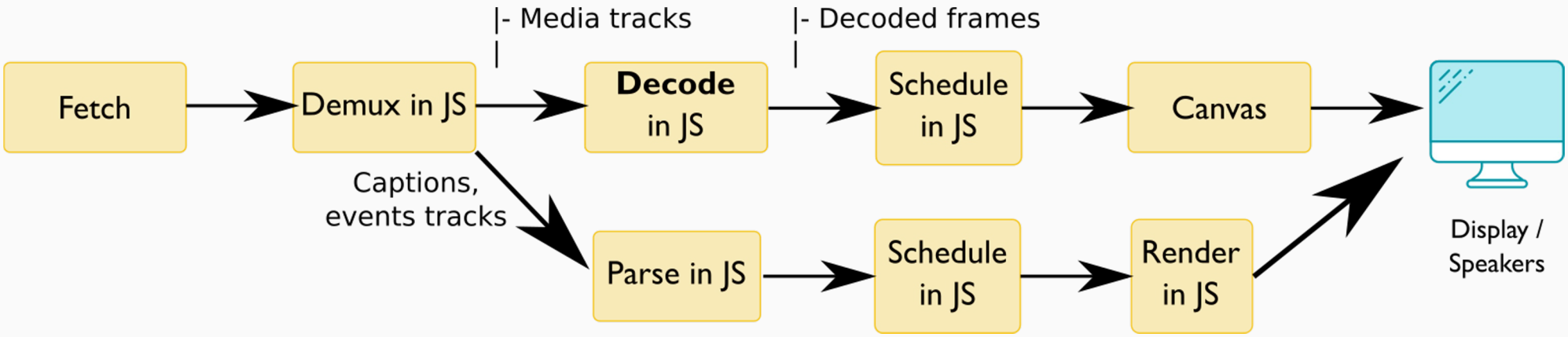
- **Live events streaming**
  - WebRTC
  - WebTransport
  - WebCodecs
- **Personalized / Engaging experiences**
  - WebCodecs & WebRTC Encoded/Decoded Media to manipulate frames and synchronize content
  - WebGPU / WebNN to process media frames

Open questions

- Exact needs?
- Performance on embedded devices?
- Is the balance between high-level and low-level primitives acceptable? Do we need an intermediary level?
- Codecs, codecs, codecs…

Media distribution in the future?
From media chunks to individual frames

- No more: <video>, MSE, WebVTT, adaptive streaming over HTTP
- Instead: WebRTC data channel or WebTransport, app processes/renders individual frames to a canvas




Web Games

Not new… First W3C workshop in 2011

- Games are **always** a driving use case
- They push platforms to their limits
- First workshop discussed HTML5 and then “new” APIs: Gamepad, Fullscreen, Pointer lock, Workers, WebRTC, WebGL, Web Audio

https://www.w3.org/2011/09/games/

Second W3C workshop in 2019

W3C Workshop on Web Games

27-28 June 2019; Redmond, WA, USA
Translations: Chinese

https://www.w3.org/2018/12/games-workshop/report.html

- Cloud gaming: WebCodecs, WebTransport
- Advanced rendering & computation: Web audio, WebGPU, WASM, threads
- Latency: inputs latency, assets loading & storage
- Discoverability & monetization

Web games: Strengths & weaknesses

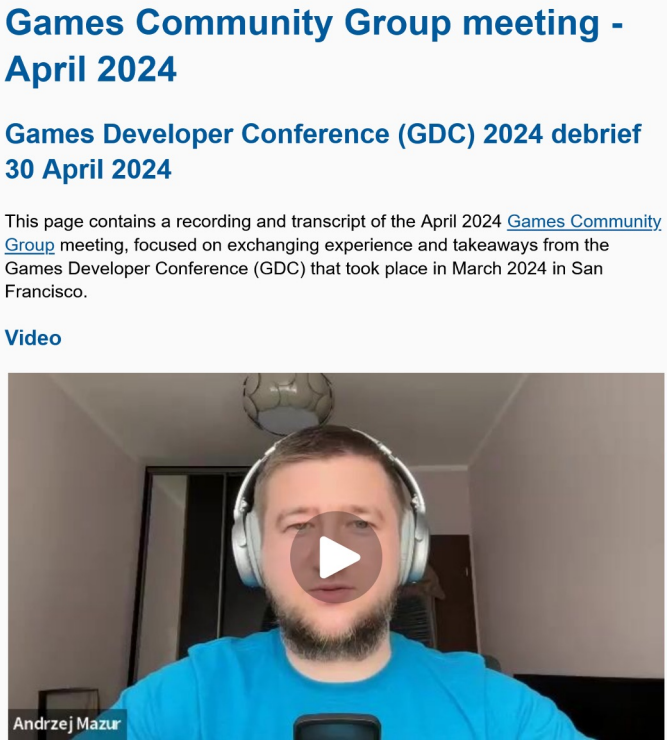
- 💪 No install step
- 💪 Secure by default
- 💪 Privacy friendly - user is anonymous
- 💪 Choice of distribution stores
- 💪 Choice of monetization scheme
- 💪 Choice of game engine, platform SDKs
- 💪 Network support included

- 🙀 Loading times, storage
- 🙀 Mixing content (e.g., ads & COEP/COOP)
- 🙀 Authentication introduces friction
- 🙀 Discoverability is hard
- 🙀 No easy way to monetize content
- 🙀 Fragmentation
- 🙀 No raw access to TCP/IP, hard to do peer-to-peer, etc.

The Games Community Group

- Created in 2011 and still active!
- **Recent discussion:**
  Regain of enthusiasm for the web as a platform at the Games Developer Conference (GDC):
  https://www.w3.org/2024/04/gamescg-gdc.html

  See also post on Web export from Godot engineer:
  https://godotengine.org/article/progress-report-web-export-in-4-3/
- **Next:** bring game platforms together and discuss alignment on web games packaging and SDKs




Web & AI

Context - ML on the Web

- W3C Workshop on Web and Machine Learning in 2020
  https://www.w3.org/2020/06/machine-learning-workshop/report.html
- Led to work on the WebNN API:
  https://www.w3.org/TR/webnn/
- ML models surfaced by other APIs:
  - Web Speech API
  - Accelerated Shape Detection API
  - Background blur, face detection, gaze correction controls in Media capture

Context - ongoing AI revolution

- Large Language Models / Generative AI
  - Trained on content crawled from the Web
  - Generate content… at scale
- Conceptual issues
  - Biases (no training is perfect)
  - Artificial hallucinations
- Scaling issues
  - Content generated in quantities beyond that developed by humans
  - No way to review/curate all content
  - Energy consumption
- Privacy issues
  - Risk of training on private data
  - Impersonation

Prompt: “An AI, hallucinating.”
Result: Exploding human head. Really?

AI & the Web report

- AI & the Web: Understanding and managing the impact of Machine Learning models on the Web
  https://www.w3.org/reports/ai-web-impact/
- Published in May 2024
- Analysis of the systemic impact of AI systems
- Captures the current shared understanding of W3C Team

Possible standardization areas

- A **consent mechanism** for the use of Web content in training pipelines
- **Labeling** content as computer-generated
- **Surfacing training sources** in model cards
- Exposing model-backed Web APIs
- **Personal data stores** to reduce risk of private data exposure
- Strengthening **credentials** and **identity** mechanisms in light of new impersonation risks
- An evaluation framework for the **environmental** impact of Web standards
- A framework to manage **interoperability** based on model inference, including for non-deterministic models

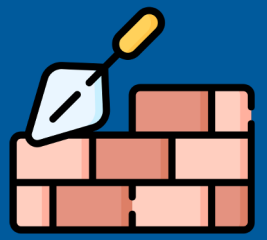
The metaverse

Source: Fortnite’s Travis $cott concert

The promised wand

- Social interactions
  - Creativity boost 🎭
  - Inclusiveness 🏳️‍🌈🐐
- Natural interactions
  - Gestures 👋👍
  - Voice 🦜
- One metaverse
  - Any device 👓📱💻✨
  - URLs FTW 🌐
- Physical & virtual
  - How many people to change a bulb in the metaverse? 💡
    Only two, you and your digital twin 👫
- Safety at all levels
  - Privacy enforced 🔒
  - Anonymity preserved whenever possible 🥷
- Distance no longer matters
  - No more business trips 💚🌎

Technical foundations **on the web**



Needed

- Some way to render 3D
- Spatial audio
- VR/AR devices support
- (Lots of) Computing power
- Real-time communications

On the web

- WebGL, WebGPU, WGSL
- Web Audio API
- WebXR set of specifications
- WebGPU, WebNN, WebAssembly
- WebRTC

Other needs?

- Decentralized
- Permissionless
- Fully functioning economy
- Digital twins
- Social presence

Do we need to worry about these in this context?

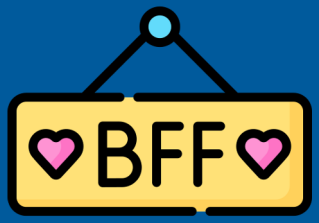
Technical needs beyond foundations

- 3D objects
- Avatar description
- Animations and their transmission in real time
- Volumetric video
- Affordances and user interactions
- World physics and defaults
- Level of Detail (LoD) and incremental streaming
- Handling of transient network failures (UX)
- …

Hurdles along the way

- Platforms create silos
  - Restricted set of proprietary platforms 💰
  - Restricted set of authorized devices 👓
  - No way to share content 🔒
- … could know everything about you
  - Your moves, where you look, what you say 👮
- … all the time
  - Always on ☀️
- … insisting on social experience
  - Anonymity, what for? Bye, privacy! 🪦
- … powered by ads
  - Dreadful user experience 🙀
- … and inclusiveness is a vague concept
  - Accessibility? On the TODO list, promised! ♿
  - Police everywhere means openness, right? 👮

The web & the metaverse

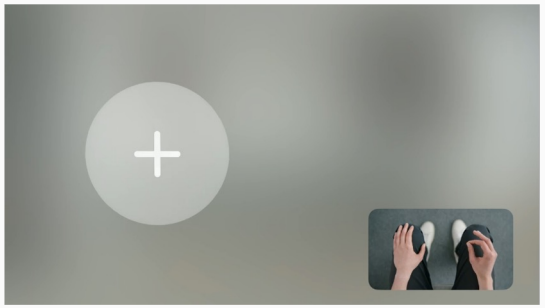


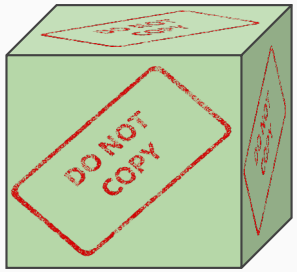
I view the metaverse as
an **evolution of the web**.

To be successful, it needs to
follow the **same core principles**!

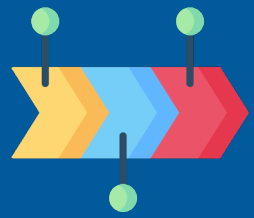
Core web principles

- Accessibility
- Internationalization
- Privacy
- Security
- Ubiquity





The web loves 1D



- The “T” in HTML stands for **Text**.
- The DOM is a tree, but main content is **linear text**.
- CSS lays out boxes **linearly** by default, one after the other.
  Even CSS Flexbox is 1D.

“Provide **text alternatives** for any **non-text content** so that it can be changed into other forms people need, such as large print, braille, **speech**, symbols or simpler language”
WCAG 2.2 - Guideline 1.1 - Text Alternatives

The web is ok with 2D

… with text alternatives!

- **Tables**
  - Various accessibility-related features to improve readability by assistive tools
  - CSS improvements have fixed abuses of tables for layout
- **Images**
  - Raster images require text alternative (except for purely decorative images!)
  - Scalable Vector Graphics (SVG) provides more semantic opportunities
- **Videos**
  - Captions and transcripts
- **Canvas**
  - Pixel control to the application!
  - Great enabler… but fully opaque!
  - Misused to design User Interfaces?

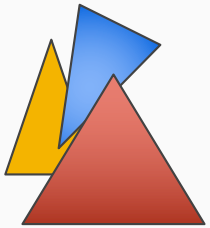
The web does not understand 3D

- <canvas> - draw your own pixels!
- We'll help, GPU APIs to draw triangles:
  - WebGL
  - WebGPU / WebGPU Shading Language

Some areas that are not covered:

- 3D asset/texture/scene formats
- User interaction
- Text 😯

How to draw a 3D owl on the web



1. Draw some triangles

2. Draw the rest of the owl

XR experiences on the web

- 2 <canvas> - one per eye
- + Device APIs to track the user's pose and gestures
  - WebXR Device API
  - WebXR Gamepad Module
  - WebXR Hand Input Module
- + AR features for blending with the physical world
  - WebXR Augmented Reality Module
  - WebXR Depth Sensing Module
  - WebXR Hit Test Module
  - WebXR Lightning Estimation Module
  - WebXR Anchor Module
  - WebXR Raw Camera Access Module
- What about accessibility?

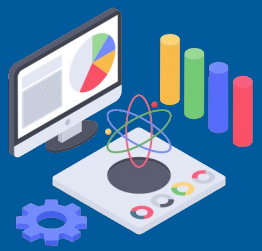
Is there a 3D user agent in the room?

Web browsers are **2D user agents**.

For the web to morph into the metaverse,
user agents need to morph into **3D user agents**.

They do not necessarily need/want that for now…
but new devices may change the *statu quo*.

Towards “2.5D” user agents



- Make 3D a first class citizen on the Web
  - The `<model>` element proposal
- Render “regular” web content in XR
  - WebXR DOM Overlays Module

Converge on 3D format(s)?

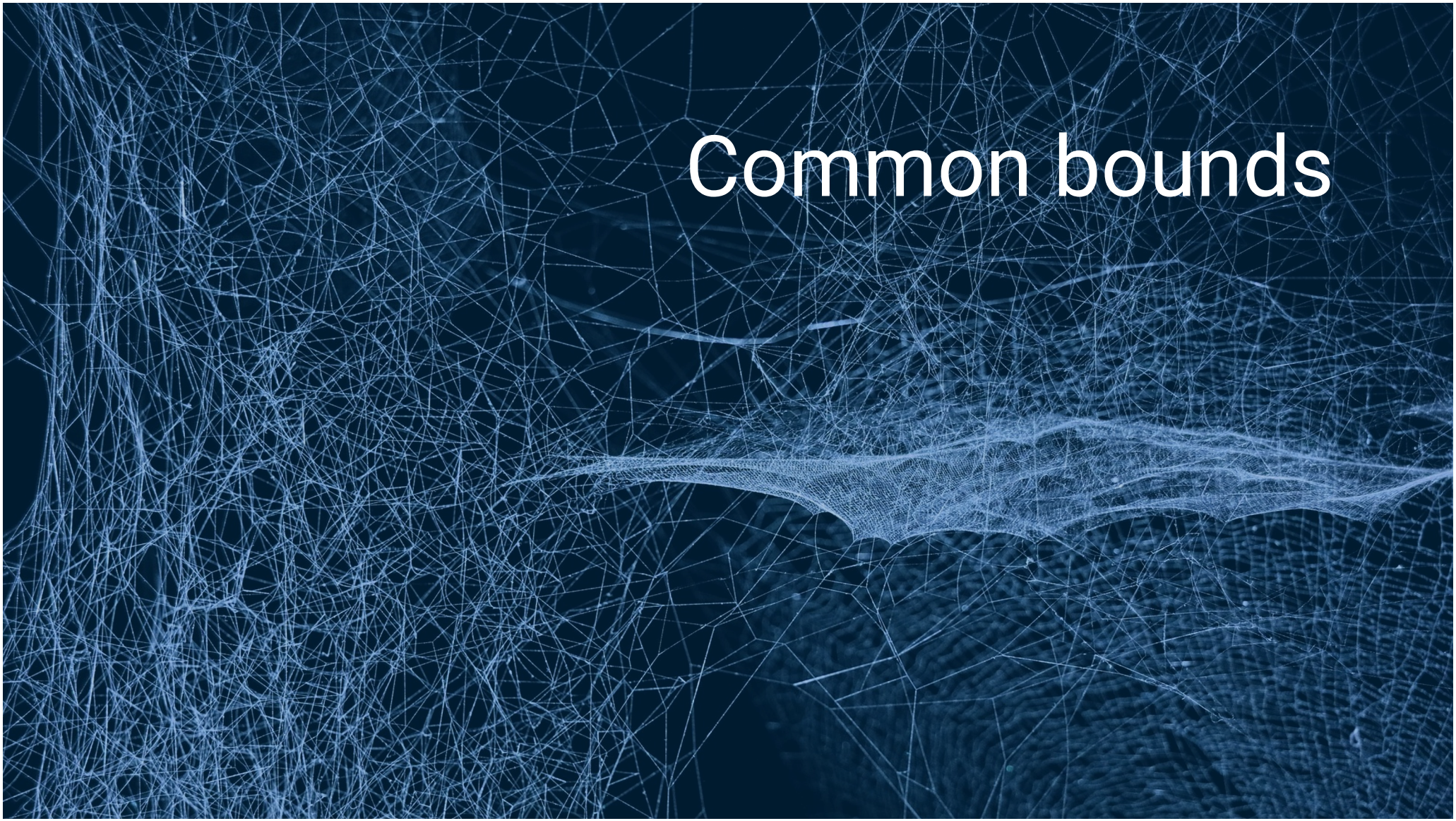
- Integrate discussed principles as much as practical
- **Industry support**
- Royalty-free
- Describe more than pixels
  - Needed for accessibility / internationalization / privacy / security / ubiquity
  - Semantics need to be as ingrained as possible
- Include more than just geometry in any case
  - Animations
  - Lights
  - Level of Detail and incremental streaming
- Avatar representation?
  - Taxonomy to describe facial expressions, gait

Interaction, locomotion, navigation

- User interface affordances
- Default world physics and dimensions
- Privacy-friendly user interaction paradigms
- Safe locomotion mechanisms
  - How to convey these mechanisms to users?
- Find appropriate navigation boundaries
  - Security
  - Continuity of experience
- Explore group experiences
  - Identity
  - Group navigation

Describe the One Web vision

- What is the One Web vision for immersive experiences?
- Progressive enhancement


Common bounds

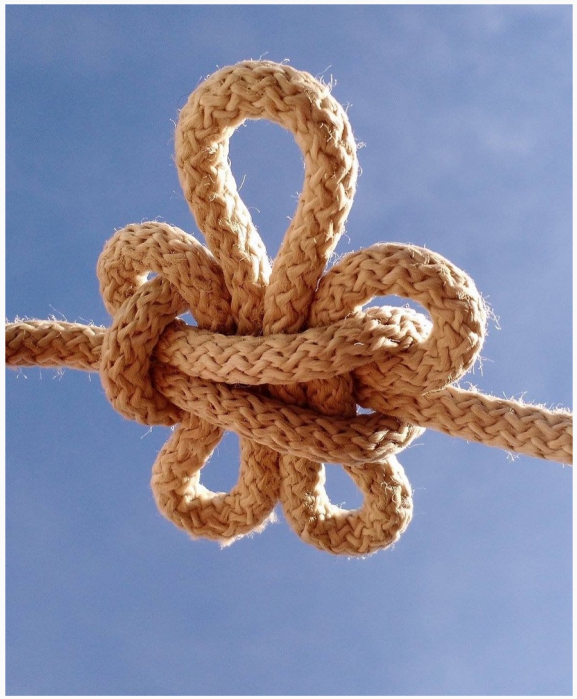
Same move towards lower-level primitives

Lower level and closer to hardware:

- WebAssembly - “native CPU”
- WebGPU - GPUs
- WebCodecs - encoders decoders
- WebNN - GPUs/NPUs/TPUs
- WebXR - Headsets
- WebRTC / WebTransport - raw network access

Growing complexity everywhere

- Intermediary library level required
  E.g., dash.js, three.js, Babylon.js, TensorFlow.js
- APIs balance simplicity vs. completeness
  E.g., `importExternalTexture()` in WebGPU
- Frustration! Always features that APIs do not readily support, e.g., hardware codecs
- Experts needed



Same need for mixing technologies

WebRTC
+
WebCodecs
+
WebNN
+
WebAssembly
+
WebGPU
+
…

x Workers

x Streams

x …

Same need to follow core web principles

- Accessibility
- Internationalization
- Privacy
- Security
- Ubiquity

Lots of activity… and long term goals

Thank you!

François Daoust
fd@w3.org
@tidoust

Attributions

Icons from www.flaticon.com:

- Virtual reality icons created by Nikita Golubev
- Secret agent, Brickwall, Question mark, Must have, Medal, Website, BFF, Cube, VR, Hard work, Timeline, 2D icons created by Freepik
- Promise icons created by Eucalyp
- Computer icons created by vectorsmarket15
- Wish list icons created by monkik
- Images from pixabay, including from garten-gg, Marco Garcia, ddzphoto