The Web of Worlds – Towards a Referential Architecture for Open Spatial Worlds

Holger Graf°
Fraunhofer Institute for Computer Graphics

Johannes Behr+
Threedy GmbH.

Alain Pagani–
DFKI GmbH

Tracey Swales‡
Portable Kingdom, Ltd.

Nicholas Polys‡‡
Virigina Tech



ABSTRACT

This paper formalises the Web of Worlds (WoW) initiative as a Web-native architectural framework for cross-world XR experiences that enables 3D spatial content to break free from existing technological silos, building on the Metaverse Standards Forum's (MSF) 3D Web Interoperability work. The presented work derives requirements from the current research initiatives on interoperability frameworks for XR and the requirements set out within the EU Virtual Worlds partnership and the MSF, by showing how existing web standards and technologies might be able to deliver open spatial worlds and interactive experiences across decentralised platforms. The position paper presents a referential 3D web architecture specifying an Open Spatial Worlds (OSW) API, a lightweight glue layer that exposes linked spatial assets and linked spatial experiences over existing Web and 3D formats (X3D, glTF, USD, OGC APIs). The paper further introduces the "Digital You" manifest and 3D Web user agent as a user-centred runtime for portable identity, consent and accessibility, and define an interoperable AI endpoint model based on MCP for spatial perception and agentic assistance. An initial implementation in industrial digital-twin pilots demonstrates already a basic cross-engine interoperability and runtime performance, providing the XWorlds community with a concrete, standards-aligned basis for Web 4.0 virtual worlds. We discuss limitations and open research questions for large-scale deployment.

Index terms: Virtual Worlds, Web4.0 Reference Architecture, Spatial Interoperability, Digital You

1 INTRODUCTION

Virtual worlds are becoming a strategic component of digital transformation, providing persistent and immersive environments for work, learning, collaboration and service access across industrial, societal and public domains. The European Commission's Web 4.0 and Virtual Worlds strategy [1] recognises these environments as enablers for industrial competitiveness, human centric services and new forms of social participation. At the same time, XR standardisation efforts reveal an ecosystem that is broad, scattered and confusing today.

Runde's global survey of XR norms [2] identifies more than 40 standards organisations and over 600 XR related documents, with a strong concentration on formats, coding and communication, but far less coverage of interaction, scene linking, identity portability and application governance. This fragmentation is reflected in virtual worlds practice: most experiences are built as vertically integrated "microverses" tightly coupled to a specific engine, identity system and distribution platform. Movement between worlds is typically mediated by 2D Web pages and login flows rather than seamless cross world transitions [3].

From a Web perspective, however, many required building blocks already exist. The Open Web Platform [4] provides URI based addressing, secure transport and browser based runtimes. WebXR and WebGPU extend browser capabilities towards immersive rendering and client side performance. X3D and glTF offer rich scene graph and asset formats [6, 7]. Decentralised Identifiers (DIDs) and Verifiable Credentials (VCs) [8] provide a basis for portable, privacy preserving identity. Recent 3D Web interoperability work within the Metaverse Standards Forum [9] shows that these ingredients can support open metaverse scenarios, but also highlights the absence of a coherent architectural layer that turns them into an operational stack.

This paper addresses that missing layer by formalising the Web of Worlds (WoW) initiative as a Web native architectural framework for cross world XR. Aim is to enable "XWorld"-scenarios in which users and AI agents can traverse, compose and enrich heterogeneous worlds across engines and devices, with continuity of identity, interaction and trust. Our contributions are:

1. A requirements analysis for Web native spatial interoperability, derived from a XR standards surveys [1], MSF use cases [9] and the EU Virtual Worlds SRIA [10].
2. A WoW reference architecture enabled by an Open Spatial Worlds (OSW) API, a lightweight glue layer linking spatial assets and experiences across existing formats and platforms.
3. A Digital You manifest and 3D Web user agent model enabling portable, user controlled identity, consent and accessibility across worlds.
4. An interoperable AI endpoint framework, based on MCP, for spatial perception and agentic assistance in XWorlds scenarios.

° e-mail: holger.graf@igd.fraunhofer.de
+ e-mail: johannes.behr@threedy.io
- e-mail: alain.pagani@dfki.de
‡ e-mail: tracey.swales@portablekingdom.co.uk
‡‡ e-mail: npolys@vt.edu

5. Initial implementation evidence from industrial Digital Twin (DT) pilots, together with open research questions for large scale deployment.

2 Building a Web of Worlds: Gaps & Challenges

Web3D research has shown for more than two decades that shared interactive 3D environments at Web scale are technically feasible [3, 7, 8, 11]. The longitudinal evolution of VRML and X3D [7] and the emergence of WebXR and WebGPU [8] demonstrate that rich immersive experiences can be delivered consistently to any connected user across platforms. Yet, as Havele et al. [3] argue, the main obstacles to an open, interoperable metaverse are cultural and economic as much as technical: poor user experience, poor corporate cooperation, and closed ecosystems that create lock in. [2] reveals that most XR standardisation activity concentrates on formats/coding and graphics software, while far fewer documents address interaction, mapping, or application level semantics. This imbalance leads to several structural gaps when one attempts to build an open Web of Worlds.

Fragmented spatial assets. Today's 3D assets are distributed across multiple formats (glTF, X3D, USD, JT, STEP, 3MF, etc.), with different capabilities for linking, hierarchy and semantics. IANA's model/* content type registry lists more than 20 registered 3D formats, each with distinct strengths [12, 13]. While scene graph standards like X3D can link external sub scenes via Inline nodes, many popular asset formats (e.g. glTF, OBJ, STL) have no native notion of cross asset linking. As a result, assets cannot be addressed, shared and reused in a uniform way through Web mechanisms.

Isolated spatial experiences. Virtual worlds are commonly deployed as monolithic applications or platform bound spaces, with little capacity for linking entire experiences or specific viewpoints using Web addresses. Users cannot easily bookmark a stateful location in an immersive world and share it with others [3]. Authentication and encryption are handled by platform specific stacks rather than standard Web protocols.

Non portable identity and consent. Identity is typically platform specific, with separate accounts, avatars, preferences and terms for each microverse. There is no widely accepted mechanism for carrying a user's avatar, preferences, accessibility needs, credentials and consent flags across heterogeneous worlds in a user controlled way. From a regulatory perspective, this is misaligned with GDPR's emphasis on user agency and data portability, and with the emerging eIDAS 2.0 and, e.g. the EU Digital Identity Wallet ecosystem.

Opaque AI integrations. AI capabilities in virtual worlds are usually embedded as opaque platform features – NPC behaviours, recommendation engines, or content generation – without standardised interfaces. AI services are rarely exposed as composable endpoints that can be discovered, orchestrated and invoked across different worlds, engines and infrastructures.

The EU Virtual Worlds SRIA [10] further stresses cross cutting challenges such as human rights compliance, safety, sustainability, data governance and trust. These requirements cannot be met by a single platform, but demand an open, interoperable architectural framework at the Web level.

3 Requirements for a Web Native Spatial Layer

Starting from the MSF use cases [5], XR standards gaps [2] and SRIA analysis [10], we derive functional requirements in four areas.

3.1 Linked Spatial Assets

A single URI must reference a discrete asset instance, which may itself be a composition of models and scenes (e.g. an X3D Inline node that recursively loads external content). Users must be able to:

- store and bookmark such URIs for later use,
- share them with other users, and
- resolve associated Web resources, including spatialised audio.

Asset instances must support linking to any number of additional instances via URIs, forming graphs of interconnected assets. Critically, the addressing layer must remain format agnostic: multiple 3D encodings (X3D, glTF, USD, JT, STEP, etc.) can be used, identified through their IANA Content Type, without affecting the URI semantics.

3.2 Linked Spatial Experiences

Beyond individual assets, the architecture must support linking complete spatial experiences. A single URI should enable a user to share and navigate to a discrete 3D experience—analogous to X3D Anchor semantics, but extended to stateful, immersive contexts. This implies:

- stable, addressable world instances,
- addressable viewpoints within worlds; and
- robust handling of authentication and encryption, leveraging existing Web stacks (HTTPS, TLS, OAuth) rather than bespoke mechanisms.

These capabilities underpin XWorld scenarios where users can jump directly from one mixed reality experience to another, or rejoin a previously visited situation with context intact.

3.3 Efficient Delivery of Spatial Experiences

Large spatial datasets may live on remote servers or distributed clouds as authenticated services. The architecture must support:

- progressive streaming, compression and chunked transfer of multi format 3D payloads,
- data driven customisation of scenes (e.g. highlighting components based on user profile),
- digital signatures to ensure asset authenticity, and
- encryption to protect sensitive content.

All of this must be handled using standard Web infrastructure (HTTP/S, CDNs) to minimise friction and cost.

3.4 User Data Manifest and Digital Presence Consistency

Users must be able to travel to and between virtual worlds, carrying a manifest of personal data – avatar, apparel, preferences, accessibility needs, consent flags, payment handles, NFTs and Verifiable Credentials – governed by three principles:

- Ownership: users retain self sovereign control over their data (DIDs, SSI).
- Permissions: users decide who can access which data and when (privacy by default).
- Decentralisation: storage is not tied to a single authority, but distributed.

A simple, standard based format (e.g. JSON LD) should encode this manifest, enabling selective disclosure through DID enabled digital wallets that provide Verifiable Presentations and zero knowledge proofs when users visit virtual worlds.

AI agents must follow the same normative framework. Non-human actors should be represented through structured metadata, authenticated, and governed by the same access and consent rules that apply to human users, enabling symmetric treatment under governance policies.

4 The Web of Worlds Reference Architecture

To meet these requirements, we propose the WoW reference architecture: a layered and open design that builds on Open Web

Platform principles while introducing a minimal glue layer for spatial interoperability (ref. figure 1).

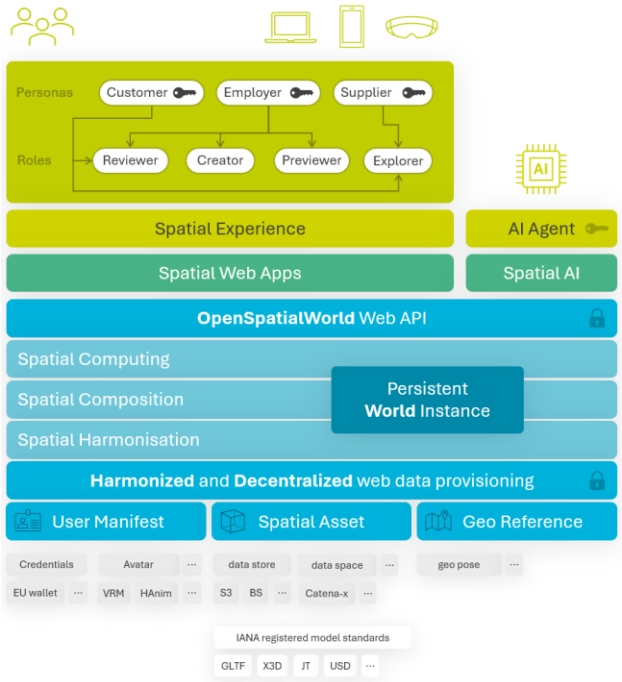


Figure 1: Layered Design of the Reference Architecture for the WoW and OSW as an instance of it

At its core is the Open Spatial Worlds API, a lightweight architectural layer that enables linking, composition and orchestration of Virtual Worlds on top of the Open Web Platform.

- linked spatial assets and fragments,
- linked spatial experiences and viewpoints,
- Digital You manifests and consent flows, and
- AI/MCP endpoints for spatial computing.

Underneath, a decentralised Web data layer resolves identifiers, retrieves and translates data states, and traverses linked structures across distributed infrastructures.

4.1 Linked Spatial Layer and Resolver Framework

A key mechanism is the introduction of linked spatial assets and linked spatial experiences as web-native resources. Assets, fragments, scenes and viewpoints become addressable, bookmarkable, composable and traversable across providers and engines through standard web mechanisms. The *Linked Spatial Layer* introduces two fundamental resource types:

- SpatialAsset: addressable 3D assets and fragments with URIs, content type metadata, optional semantic annotations, and optional links to other assets.
- SpatialWorld: addressable world instances with URIs, spatial composition graphs, portals to other worlds, and references to views and users.

A harmonised resolver framework maps these URIs to data states through four stages (ref. figure 2):

1. Identification: interpret the URI (URL, DID, URN) and determine applicable resolution methods.
2. Retrieval: obtain the raw payload (3D asset, scene graph, manifest) via HTTP/S or DID resolution.
3. Translation: convert payload into domain specific structures (scene graphs, spatial composition nodes), extracting embedded URIs.
4. Traversal: follow linked URIs to discover further assets, worlds and user manifests.

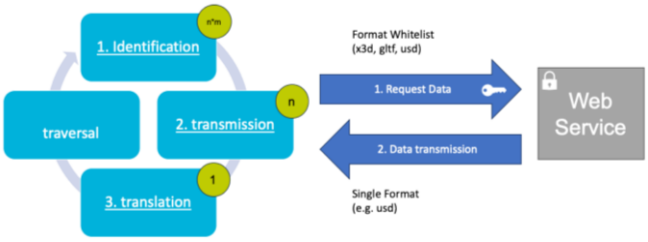


Figure 2: WoW asset data provisioning based on the resolver framework

This process extends Web3D's longstanding Anchor and Inline patterns [4, 11] into a more general, format agnostic linking layer that works across X3D, glTF, USD and other formats registered in RFC 2077 and the IANA model/* registry [12, 13].

4.2 Spatial Composition Graph

Most virtual worlds comprise many assets, not a single static file. We model this through a dynamic Spatial Composition Graph: a hierarchical structure in which nodes reference assets or subgraphs, with local transformations and child relationships. Nodes may be:

- linked to external SpatialAssets via URIs;
- linked to internal nodes within the same world; or
- linked to nodes in other worlds via portals.

This enables XWorlds scenarios where a world can inline a portion of another world, expose portals to external worlds, or share portions of its composition graph across devices and sessions.

4.3 Digital You Manifest

The Digital You manifest will be specified as a JSON LD document anchored to a DID. It comprises:

- avatar and apparel references,
- preferences and accessibility settings,
- consent flags and privacy metadata,
- payment handles (fiat/crypto),
- NFT pointers and digital asset references, and
- Verifiable Credential pointers or inline VCs.

It is carried through DID enabled wallets, which store private keys and credentials and construct Verifiable Presentations or zero knowledge proofs upon request. Worlds request only the minimum necessary subset of manifest data; wallets enforce user consent and selective disclosure.

This design is aligned with SSI principles [9], GDPR and eIDAS, and provide a human centric basis for portable identity and digital presence consistency, as advocated by Havele et al. [5] and the EU Web 4.0 strategy [1].

4.4 3D Web User Agent

The 3D Web user agent is the runtime concept through which users interact with the WoW architecture. Conceptually analogous to the browser for the 2D web, it is responsible for:

- discovering and loading SpatialWorlds via URIs,
- resolving and composing spatial assets via the OSW API,
- mediating Digital You manifest exchanges and consent prompts,
- enforcing security and privacy policies, and
- providing device adapted interaction via WebXR and WebGPU.

Rather than defining a single implementation, WoW treats the user agent as a multi target runtime concept that can be realised as browser extensions, engine integrations or standalone shells while sharing the same OSW interfaces and conformance suite. This supports consistent behaviours across desktop browsers, mobile devices and XR headsets.

4.5 Agentic AI Layer and MCP Endpoints

AI services are integrated through an agentic AI layer using the Model Context Protocol (MCP) as a foundation. This layer comprises:

- spatial context providers exposing world state, scene graphs, user positions and environment metadata;
- spatial tools allowing AI agents to query scenes, compute distances, detect collisions, generate or adapt assets, and annotate worlds; and
- capability manifests describing each AI module's abilities and constraints.

AI agents discover and invoke these tools using MCP style interactions [14, 15], enabling spatially grounded reasoning, perception and generation. Workloads are distributed across cloud, edge and device infrastructures according to context constraints (latency, privacy, energy), as recommended in the Virtual Worlds SRIA.

5 Initial Implementation: Industrial DT-Service Worlds

To validate the architecture, an initial pilot in the industrial digital twin domain has been realised, where virtual representations of machines and production lines are already common but service workflows remain fragmented [2, 5].

In this scenario, a technician identifies a physical machine via a product ID or Digital Product Passport. This ID resolves to a SpatialWorld URI representing a Service World for that machine. Through the 3D Web user agent, the technician enters a web based immersive environment showing the machine's 3D representation (JT/STEP translated into X3D/glTF), maintenance history and live status, all orchestrated through the OSW API.

The Digital You manifest carries the technician's professional credentials, safety certifications and preferences, enabling the Service World to adapt content and access levels while enforcing least privilege data access. AI endpoints provide spatial assistance, such as computing minimum distances between parts, highlighting components due for inspection, or generating repair instructions based on twin data and context.

Preliminary measurements on this pilot demonstrate:

- successful cross engine asset reuse: machine models from CAD toolchains streamed into web based engines via OSW compatible asset services,
- scene loading times under 3 seconds on standard hardware for typical service worlds,
- AI inference latency under 500 ms for spatial queries and guidance tasks, and
- compatible Digital You manifest handling across web and XR entry points.

These early results suggest that a Web native WoW architecture can support interoperable industrial service worlds, although further performance optimisation and robustness testing are required for large scale deployment.

6 Discussion and Open Questions

The WoW architecture is deliberately minimal and compositional, seeking to reuse existing standards rather than invent new ones. Compared with monolithic frameworks like OpenUSD based Omniverse [16, 17], it emphasises URI based linking, Web native runtimes and explicit identity and AI layers. However, several open questions remain:

- Standardisation scope. How much of the WoW stack should be standardised? URI patterns for spatial assets and experiences, Digital You schemas, and AI endpoint interfaces are strong candidates, but care must be taken to avoid over specification that stifles innovation.
- Security and abuse prevention. Spatial linking and cross world traversal introduce new attack surfaces (e.g. malicious portals, asset poisoning, identity impersonation). The security framework sketched here needs to be elaborated and validated through threat modelling and empirical studies.
- Governance and moderation. WoW provides the technical basis for interoperability, but governance of content and behaviour across worlds remains a socio technical challenge. Aligning WoW mechanisms with emerging governance models for XR and virtual worlds [3, 10] is a key research topic.
- Usability and accessibility. While the architecture supports accessibility profiles and preference transport, user studies are needed to assess whether Digital You, consent flows and the 3D user agent model are understandable, manageable and inclusive across diverse populations.

Addressing these questions will require coordinated work across standards bodies, research projects and industrial deployments.

7 Conclusion

We have presented the Web of Worlds (WoW) initiative as a Web native architectural framework for cross world XR interoperability, building on the MSF 3D Web Interoperability work and the EU Web 4.0 and Virtual Worlds strategy. By analysing requirements from XR standards surveys and the Virtual Worlds SRIA, we identified gaps around linked spatial assets, linked spatial experiences, portable identity and AI integration. We defined a reference architecture and OSW API that provide a minimal glue layer across existing formats and platforms, introduced the Digital You manifest and 3D Web user agent for human centric identity and runtime control, and outlined an agentic AI layer based on MCP endpoints for spatial computing.

Initial industrial digital twin pilots demonstrate that this architecture can support cross engine interoperability and acceptable runtime performance in realistic service scenarios. While further research is needed on security, governance and usability, we believe that WoW offers an actionable, standards aligned basis for Web 4.0 virtual worlds that is open, decentralised and human centric.

References

[1] European Commission. An EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition. COM(2023) 442, July 2023.

[2] C. Runde. XR Standardization: A Global Overview. 10.6084/m9.figshare.24752118., 2023.

[3] J. Behr, A. Havele, N. Polys: Building 3DWeb Interoperability for the Metaverse, Web3D '23: Proceedings of the 28th International ACM Conference on 3D Web Technology, October 2023, https://doi.org/10.1145/3611314.3623584

[4] World Wide Web Consortium. Open Web Platform. W3C Wiki. [Online]. Available: https://www.w3.org/wiki/Open_Web_Platform. [Accessed: Jun. 29, 2026].

[5] J. Behr, T. Swales, N. Polys: The Web of Worlds (WoW) Whitepaper - Enabling an open, decentralized and efficient Metaverse on the Open Web Platform. In: Metaverse Standard Forum, 3D Web Interoperability Domain Group Release: Q1, draft: 2026-03-31, 2026. https://webofworlds.github.io/initial_MSF_Whitepaper/gen/MSF-3DWebInterop_WoWWhitepaper.pdf

[6] A. Rodríguez, P. L. García, J. S. Marquerie, R. Marques, J. Blat: A Cross-Platform, WebGPU-Based 3D Engine for Real-Time Rendering and XR Applications, In Web3D '25: Proceedings of the 30th International Conference on 3D Web Technology, 2025, https://doi.org/10.1145/3746237.3746305

[7] A. Havele, N. Polys, W. Benman, and D. Brutzman: The Keys to an Open, Interoperable Metaverse. In Proceedings of the 27th International Conference on 3D Web Technology (Web3D '22). Association for Computing Machinery, New York, NY, USA, pp. 1–7, 2022. https://doi.org/10.1145/3564533.3564575

[8] C. Mazzocca, A. Acar, S. Uluagac, R. Montanari, P. Bellavista and M. Conti: A Survey on Decentralized Identifiers and Verifiable Credentials, in IEEE Communications Surveys & Tutorials, vol. 27, no. 6, pp. 3641-3671, 2025, doi: 10.1109/COMST.2025.3543197.

[9] N. Polys, J. Behr, T. Swales: The Web of Worlds (WoW) Whitepaper - Enabling an open, decentralized and efficient Metaverse on the Open Web Platform. In: Metaverse Standard Forum, 3D Web Interoperability Domain Group Release: Q1, draft: 2026-03-31, 2026. https://webofworlds.github.io/initial_MSF_Whitepaper/gen/MSF-3DWebInterop_WoWWhitepaper.pdf

[10] Virtual Worlds Association. (2025). Strategic research and innovation agenda (SRIA) for virtual worlds. European Commission Horizon Europe. https://www.virtualworldsassociation.eu/actions/strategic-research-innovation-agenda-virtual-worlds-eu [Last Accessed: 11.07.2026]

[11] J. Behr, C. Mouton, S. Parfouru, J. Champeau, C. Jeulin, M. Thöner, C. Stein, et al. WebVis/Instant3DHub: Visual Computing as a Service Infrastructure to Deliver Adaptive, Secure and Scalable User Centric Data Visualisation, 2015. https://doi.org/10.1145/2775292.2775299

[12] S. Nelson, C. Parks and Mitra: The model primary content type for multipurpose internet mail extensions (RFC 2077). IETF. https://doi.org/10.17487/RFC2077

[13] IANA media types registry:https://www.iana.org/assignments/media-types/media-types.xhtml#model [Last Accessed: 05.07.2026]

[14] S. Wu and J. P. Barnett: MCP-Unity: Protocol-Driven Framework for Interactive 3D Authoring. In Proceedings of the SIGGRAPH Asia 2025 Technical Communications (SA Technical Communications '25). Association for Computing Machinery, New York, NY, USA, Article 7, 1–4, 2025 https://doi.org/10.1145/3757376.3771417

[15] Partha Pratim Ray. A Survey on Model Context Protocol: Architecture, State-of-the-art, Challenges and Future Directions. *TechRxiv*. 18 April 2025. DOI: https://doi.org/10.36227/techrxiv.174495492.22752319/v1

[16] Ahmed, N. et al. A systemic survey of the Omniverse platform and its applications in data generation, simulation and metaverse. Frontiers in Computer Science, 2024

[17] Yang, L et al. Interoperability of the Metaverse: A Digital Ecosystem Perspective Review. IEEE Engineering Management Review, 2025