ETSI GR ARF 007 V1.1.1 (2025-03)

Augmented Reality Framework (ARF); Standards landscape for ETSI AR Functional Reference Model

Disclaimer

The present document has been produced and approved by the Augmented Reality Framework (ARF) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference

DGR/ARF-007

Keywords

augmented reality, interoperability

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

BLUETOOTH® is a trademark registered and owned by Bluetooth SIG, Inc.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Augmented Reality Framework (ARF).

The ISG ARF shares the following understanding for Augmented Reality: Augmented Reality (AR) is the ability to mix in real-time spatially-registered digital content with the real world. The present report provides a snapshot of standardization efforts conducted by various Standards Development Organizations (SDOs) and other fora, as available at the time of publishing. It positions the existing and emerging standards in the context of the ETSI GS ARF 003 [i.1].

While the goal of the present document is to provide an exhaustive list of relevant standards pertaining to the functional blocks and Reference Points (RPs) of the AR Framework Reference Architecture, this may not be the case with the first version of the present document and updates may be available in future versions.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document identifies existing and emerging standards and technologies relevant to the ten functional blocks and nineteen reference points specified in ETSI GS ARF 003 [i.1]. It describes how each standard contributes to addressing the interoperability needs of components, systems and services seeking to fulfil the roles of each defined block and reference point.

The present document also identifies where new or additional coordination between existing standards bodies and working groups may be needed to clarify how each contributes to any interoperability gaps and to reduce confusion emerging from standards scope overlap or conflicts.

While some of these standards included in the present document directly address an AR function, others are addressing key technological components that can be useful to increase interoperability of AR components, systems and services and their interactions with other information technologies.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI GS ARF 003: "Augmented Reality Framework (ARF) AR framework architecture".

[i.2] W3C® Recommendation 08 October 2024: "WebRTC: Real-Time Communication in Browsers".

[i.3] W3C® Recommendation 16 August 2011: "Scalable Vector Graphics (SVG) 1.1".

[i.4] W3C® Working Draft 14 February 2025: "Media Capabilities".

[i.5] W3C® Working Draft 04 July 2024: "Media Source Extensions™".

[i.6] W3C® Editor's Draft 04 July 2024: "Media Playback Quality".

[i.7] W3C® Recommendation 9 April 2020 (Link errors corrected 23 June 2020): "Web of Things (WoT) Thing Description".

[i.8] IETF RFC 9293: "Transmission Control Protocol (TCP)".

[i.9] IETF RFC 768: "User Datagram Protocol".

[i.10] IETF RFC 791: "Internet Protocol ".

[i.11] IETF RFC 8200: "Internet Protocol, Version 6 (IPv6) Specification".

[i.12] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".

[i.13] IETF RFC 7826: "Real-Time Streaming Protocol Version 2.0".

[i.14] IETF RFC 3261: "SIP: Session Initiation Protocol".

[i.15] IETF RFC 2974: "Session Announcement Protocol".

[i.16] IETF RFC 2608: "Service Location Protocol, Version 2".

[i.17] IETF RFC 6763: "DNS-Based Service Discovery".

[i.18] IETF RFC 6733: "Diameter Base Protocol".

[i.19] IETF RFC 8866: "SDP: Session Description Protocol".

[i.20] ISO/IEC 14496-3:2019:"Information technology — Coding of audio-visual objects — Part 3: Audio".

[i.21] ISO/IEC 23008-3:2022: "Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio".

[i.22] ISO/IEC 23090-5:2023:"Information technology — Coded representation of immersive media — Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC)".

[i.23] ISO/IEC 18000-3:2010: "Information technology — Radio frequency identification for item management — Part 3: Parameters for air interface communications at 13,56 MHz".

[i.24] ISO/IEC 15961-1:2021: "Information technology — Data protocol for radio frequency identification (RFID) for item management — Part 1: Application interface".

[i.25] ISO 19156:2023: "Geographic information — Observations, measurements and samples".

[i.26] ISO/IEC 23005-5:2019: "Information technology — Media context and control — Part 5: Data formats for interaction devices".

[i.27] ISO/IEC 12113:2022: "Information technology — Runtime 3D asset delivery format — Khronos glTF™ 2.0".

[i.28] Recommendation ITU-T H.265: "High efficiency video coding".

[i.29] Recommendation ITU-T H.266: "Versatile video coding".

[i.30] Recommendation ITU-T T.81: "Information technology—Digital compression and coding of continuous-tone still images—Requirements and guidelines".

[i.31] AV1 Bitstream & Decoding Process Specification.

[i.32] United States Coast Guard - U.S. Department of Homeland Security: "GPS References".

[i.33] European Space Agency (ESA): "GALILEO".

[i.34] Bluetooth®: "Core Specifications 5.2".

[i.35] IEEE 802.11az™-2002: "IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks--Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Enhancements for Positioning".

[i.36] IEEE 802.15.4z™-2020: "IEEE Standard for Low-Rate Wireless Networks--Amendment 1: Enhanced Ultra Wideband (UWB) Physical Layers (PHYs) and Associated Ranging Techniques".

[i.37] OGC® SensorML: Model and XML Encoding Standard.

[i.38] OGC® SensorThings API Part 1: Sensing Version 1.1.

[i.39] OGC® SensorThings API Part 2 – Tasking Core.

[i.40] OASIS Standard: "MQTT Version 5.0", 7 March 2019.

[i.41] LAS Specification 1.4 - R15.

[i.42] The OpenXR™ Specification.

[i.43] Khronos® - Collada™.

[i.44] Wavefront OBJ File Format Summary.

[i.45] ISO/IEC 23005-3:2019: "Information technology — Media context and control — Part 3: Sensory information".

[i.46] Draft OGC® GeoPose 1.0 Data Exchange Standard.

[i.47] ISO/IEC 14496-11:2015: "Information technology — Coding of audio-visual objects — Part 11: Scene description and application engine".

[i.48] ISO/IEC 15948:2004: "Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification".

[i.49] ISO/IEC 23007-2:2012: "Information technology — Rich media user interfaces — Part 2: Advanced user interaction (AUI) interfaces".

[i.50] W3C® Candidate Recommendation Snapshot, 31 March 2022: "WebXR Device API".

[i.51] ISO/IEC 11179-3:2023: "Information technology — Metadata registries (MDR) — Part 3: Metamodel for registry common facilities".

[i.52] IETF RFC 5013: "The Dublin Core Metadata Element Set".

[i.53] ISO 19115-1:2014: "Geographic information — Metadata — Part 1: Fundamentals".

[i.54] ISO/TS 19139-1:2019: "Geographic information — XML schema implementation — Part 1: Encoding rules".

[i.55] ISO/IEC 19775-1:2023: "Computer graphics, image processing and environmental data representation — Extensible 3D (X3D) — Part 1: Architecture and base components ".

[i.56] ISO/IEC 23005-2:2018: "Information technology — Media context and control — Part 2: Control information".

[i.57] OGC® Augmented Reality Markup Language 2.0 (ARML 2.0).

[i.58] ISO/IEC 14772-1:1997: "Information technology — Computer graphics and image processing — The Virtual Reality Modeling Language — Part 1: Functional specification and UTF-8 encoding".

[i.59] OMI Github: "Metaverse interoperability community group (X3C), OMI glTF™ Extensions".

[i.60] USD: "Universal Scene Description (Pixar)".

[i.61] FBX: "Adaptable file format for 3D animation software".

[i.62] ISO/IEC 23000-13:2017: "Information technology — Multimedia application format (MPEG-A) — Part 13: Augmented reality application format".

[i.63] ISO/IEC 14496-16:2011: "Information technology — Coding of audio-visual objects — Part 16: Animation Framework eXtension (AFX)".

[i.64] ISO/IEC 23090-14:2023: "Information technology — Coded representation of immersive media — Part 14: Scene description".

[i.65] W3C® Candidate Recommendation Draft, 21 October 2024: "WebXR Device API".

[i.66] More to explore with ARKit 6.

[i.67] Make the world your canvas.

[i.68] Snapdragon Spaces.

[i.69] ISO/IEC 15938-13:2015: "Information technology — Multimedia content description interface — Part 13: Compact descriptors for visual search".

[i.70] ISO/IEC 15938-15:2019: "Information technology — Multimedia content description interface — Part 15: Compact descriptors for video analysis".

[i.71] ISO/IEC 15938-4:2002: "Information technology — Multimedia content description interface — Part 4: Audio".

[i.72] ISO/IEC 23090-9:2023: "Information technology — Coded representation of immersive media — Part 9: Geometry-based point cloud compression".

[i.73] ISO/IEC 23090-12:2023: "Information technology — Coded representation of immersive media — Part 12: MPEG immersive video".

[i.74] IETF RFC 9113: "HTTP/2".

[i.75] IETF RFC 6455: "The Websocket Protocol".

[i.76] IETF RFC 5531: "RPC: Remote Procedure Call Protocol Specification Version 2".

[i.77] ISO 19111:2019: "Geographic information — Referencing by coordinates".

[i.78] OGC® Abstract Specification Topic 2: Referencing by coordinates Corrigendum.

[i.79] OGC® API - Features - Part 2: Coordinate Reference Systems by Reference.

[i.80] ISO 19112:2019: "Geographic information — Spatial referencing by geographic identifiers".

[i.81] Cesium® Platform: "The platform for 3D geospatial".

[i.82] CesiumGS®: "3D Tiles".

[i.83] ISO/IEC 23090-3:2024: " Information technology — Coded representation of immersive media — Part 3: Versatile video coding".

[i.84] IETF RFC 3866: "Language Tags and Ranges in the Lightweight Directory Access Protocol (LDAP)".

[i.85] glTF™ 2.0 Interactivity Extension Specification.

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

3 Degrees of Freedom (3DoF): visual system providing rotation of the viewing position around three perpendicular axes

6 Degrees of Freedom (6DoF): visual system providing rotation of the viewing position around three perpendicular axes and allowing the transition of the viewing position along these three axes

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

2D	Two dimensions
3D	Three dimensions
3DoF	3 Degrees of Freedom
6DoF	6 Degrees of Freedom
AAC	Advanced Audio Coding
AFX	Animation Framework eXtension
API	Application Programming Interface
AR	Augmented Reality
ARAF	Augmented Reality Application Format
ARF	Augmented Reality Framework
ARML	Augmented Reality Markup Language
ASCII	American Standard Code for Information Interchange
ASPRS	American Society for Photogrammetry and Remote Sensing
AUI	Advanced User Interface
AV1	AOMedia Video 1
AVC	Advanced Video Coding
AVP	Attribute-Value Pairs
BIFS	Binary Format for Scenes
BIM	Building Information Modelling
CAD	Computer Aided Design
Collada	Collaborative Design Activity
CRS	Coordinates Reference System
DGPS	Differential Global Positioning System
DNS	Domain Name System
FBX	Filmbox
glTF	Graphics Language Transmission Format
GNSS	Global Navigation Satellite System
G-PCC	Geometry-based Point Cloud Compression
GPS	Global Positioning System
GR	Group Report
HEVC	High Efficiency Video Coding
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JPEG	Joint Photographic Experts Group
JSON	JavaScript Object Notation
LAN	Local Area Network
LAS	LASer
LiDAR	Light Detection And Ranging
MIV	MPEG Immersive Video
MPEG	Motion Picture Expert Group
MQTT	Message Queuing Telemetry Transport
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geographic Consortium
OMI	Open Metaverse Interoperability
PNG	Portable Network Graphics
RAHT	Region-Adaptive Hierarchical Transform
RFC	Requests For Comments
RFID	Radio Frequency Identification
RGB	Red Green Blue
RP	Reference Point

RTCP	RTP Control Protocol
RTP	Real Time Protocol
RTSP	Real Time Streaming Protocol
SDK	Software Development Kit
SDO	Standards Developing Organization
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SVG	Scalable Vector Graphics
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UML	Unified Modelling Language
URI	Uniform Resource Identifier
USD	Universal Scene Description
V3C	Visual Volumetric Video-based Coding
VFX	Visual Effects
V-PCC	Video-based Point Cloud Compression
VR	Virtual Reality
VRML	Virtual Reality Markup Language
VVC	Versatile Video Coding
X3D	eXtensible 3D
XML	eXtensible Markup Language
XR	eXtended Reality

4 Organization of this AR Standards Landscape

4.1 Introduction

ETSI GS ARF 003 [i.1] defines a global architecture of an AR system consisting of hardware, software and data components. In addition to the global architecture, a functional architecture was defined. The functional architecture specified by ETSI GS ARF 003 [i.1] is shown in Figure 1. Logical functions are shown as named boxes that may be nested in cases where a high-level function is composed of several subfunctions. The architecture can be used for describing stand alone or "fully" integrated AR systems as well as for describing implementations of components distributed across local and remote resources (e.g. network servers). Depending on the deployment scenario and the application requirements, multiple logical subfunctions and functions can also be combined in one deployable unit.

Functions in the architecture are connected by Reference Points (RPs). A reference point in a functional architecture is located between two non-overlapping functions and represents the interrelated interactions between those functions. Reference point implementations are interfaces that convey information between functions in a unidirectional or bidirectional way using a specified protocol.

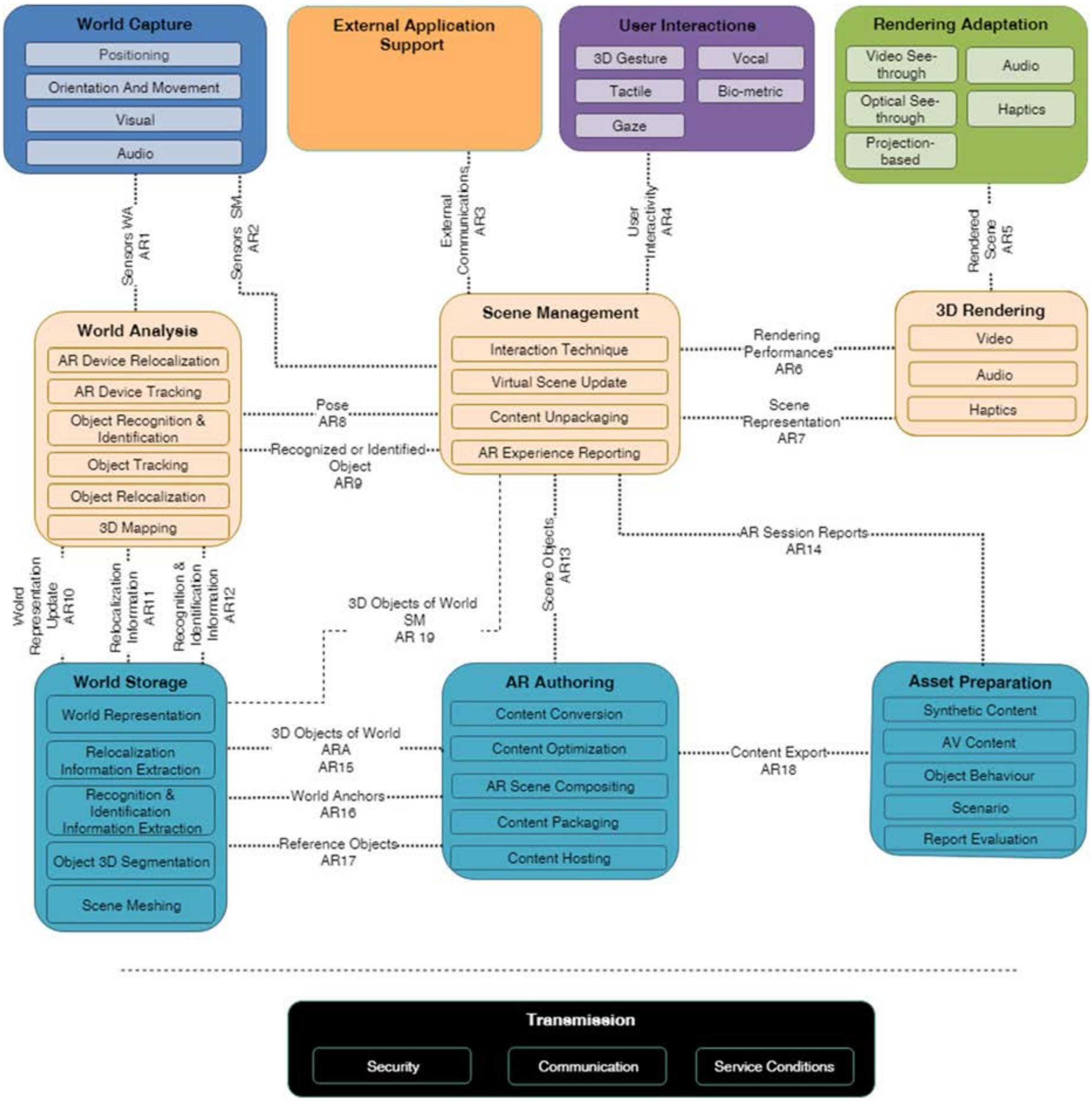

Figure 1: Diagram of the functional reference architecture

4.2 The ETSI AR Framework Functional Architecture and Associated Standards

4.2.1 Introduction

The present document describes the standards landscape and its technologies on the basis of reference points defined in ETSI GS ARF 003 [i.1] and depicted in Figure 1. For each of these RPs, the standards and technologies of potential value to AR system functions are listed in tables. These tables mention the name of the standards or technology organization and of the standard itself, further information is collected in clause 5 of the present document.

NOTE: The standards that are collected in the present document were chosen by the members and participants of the ISG as they are most commonly used in the field of AR and XR. This does not exclude the usage of other standards and their technologies, depending on application-related requirements and availability of techniques.

Taking into account that all functions shown in Figure 1 will work with IP-based technologies, the communication between them via the RPs also needs to support IP-based techniques. The functional block "Transmission" in Figure 1 takes care about this type of connectivity, its protocols and standards are listed in clause 4.2.2. From clause 4.2.3 on, relevant standards and technologies for each of the RPs are mentioned. From clause 5 on, more information about these technologies is given, consolidated under the name of the organizations that publish their specifications and take care about their maintenance.

4.2.2 Standards and technologies for IP-based Communication

For a packet-oriented communication across networks the suite of IP protocols as developed by IETF and other organizations is used. This suite of standards encapsulates data in datagrams, distributes IP addresses to source and destination and takes care about the routing of the datagrams across the network. All standards and technologies mentioned from clause 4.2.3 on make use of at least one of the protocols listed in Table 1.

Table 1: Protocols for IP-based communication

Communication Protocols	SDO	Features	Standards
Transmission Control Protocol (TCP)	IETF	Setting up a connection-oriented communication between source and destination	IETF RFC 9293 [i.8]
User Datagram Protocol (UDP)	IETF	Setting up a connectionless communication between source and destination	IETF RFC 768 [i.9]
Internet Protocol (IP)	IETF	Delivery of packets from source to destination	IETF RFC 791 [i.10] (IPv4), IETF RFC 8200 [i.11] (IPv6)
Real-time Transmission Protocol (RTP)	IETF	Delivery of audio and video from source to destination	IETF RFC 3550 [i.12]
RTP Control Protocol (RTCP)	IETF	Control information for an RTP session	IETF RFC 3550 [i.12] (RTCP, section 6 of RTP)
Real-time Streaming Protocol (RTSP)	IETF	Support for multiplexing and packetizing multimedia transport streams	IETF RFC 7826 [i.13]
WebRTC: Real Time Communication between Browsers	W3C®	Realtime communication between browsers	WebRTC 1.0 [i.2]
Session Initiation Protocol (SIP)	IETF	Protocol for setting up and controlling communication sessions	IETF RFC 3261 [i.14]

4.2.3 Standards and technologies for ARF Reference Points 1 and 2

These Reference Points support the communication between the World Capture, World Analysis and the Scene Management functions respectively to ensure that relevant features in the real world which are captured by sensors are available, received and useable by the relevant subfunctions. Sensors are used in an application-specific way and so their detection and registration to the applications happens via functionalities that are not standardized. However, for the formatting of media in the field of audio-visual content, standards are available using different compression schemes. Protocols widely used are collected in Table 2.

Table 2: Standards for the formatting and compression of audio-visual content

Protocols	SDO	Features	Standards
Advanced Audio Coding (AAC)	ISO/IEC MPEG	Audio wideband codec with several profiles as a flexible framework for audio coding, synchronization, and mixing	ISO/IEC 14496-3 [i.20]
MPEG-H 3D Audio	ISO/IEC MPEG	Audio coding for audio channels, audio objects and for full sphere surround sound.	ISO/IEC 23008-3 [i.21]
High Efficiency Video Coding (HEVC)	ITU	Video coding with several profiles for resolutions up to 8k, extension for the coding of 3D videos	Recommendation ITU-T H.265 [i.28]
Versatile Video Coding (VVC)	ITU	Successor of HEVC codec for resolutions up to 16k and support for extended colour space and bit depths up to 16 bits per component	Recommendation ITU-T H.266 [i.29]

Protocols	SDO	Features	Standards
AOMedia Video 1 (AV1)	AOMedia	Video coding for resolutions up to 16 k and bit depths up to 12 bits per component	AV1 [i.31]
Digital Image Compression JPEG	ITU	Compression scheme for digital images with an adjustable level of compression	Recommendation ITU-T T.81 [i.30]
Scalable Vector Graphics	W3C®	XML-based vector format for describing two-dimensional images	SVG 1.1 [i.3]
Portable Network Graphics (PNG)	ISO/IEC	Lossless compression scheme and file format for digital images	ISO/IEC 15948 [i.48]
Visual volumetric video-based coding; Video-based point cloud compression (V3C/V-PCC)	ISO/IEC	Compression scheme for visual 3D point clouds with the conversion of point cloud data from 3D to 2D to use existing video codecs.	ISO/IEC 23090-5 [i.22]
Visual volumetric video-based coding; MPEG Immersive Video (V3C/MIV)	ISO/IEC	Compression scheme for immersive video content, supporting several presentation schemes.	ISO/IEC 23090-12 [i.73]
Visual volumetric video-based coding; Geometry-based point cloud compression (V3C/G-PCC)	ISO/IEC	Geometry-based compression scheme for visual 3D point clouds working in 3D	ISO/IEC 23090-3 [i.83]

In the field of XR applications, the detection of the position of the device that runs the applications is important for the quality of service and for the user experience. Prominent technologies used for positioning purposes are listed in Table 3.

Table 3: Positioning technologies

Protocols	SDO / Provider	Features	Standards
GNSS: Global Positioning System GPS	United States Coast Guard	World-wide operating positioning system using satellite-based radio navigation	GPS [i.32]
GNSS: Galileo	European Space Agency	World-wide operating positioning system using satellite-based radio navigation	Galileo [i.33]
Wireless LAN Positioning	IEEE	Positioning system evaluating the propagation of radio waves	IEEE 802.11az [i.35]
Bluetooth® Low Energy Indoor Positioning System	Bluetooth®	Positioning system using Bluetooth® devices and beacons and their relations to each other	Bluetooth® Core Specification [i.34]
Ultra Wide Band Positioning	IEEE	Positioning system based on the transmission of short pulses within a wide frequency band	IEEE 802.15.4z [i.36]
Radio Frequency Identification (RFID)	ISO/IEC	Near distance location system with radio detection	ISO/IEC 18000-3 [i.23]
LAS specification 1.4	File format by American Society for Photogrammetry and Remote Sensing	Positioning system using laser pulses by scanning objects and calculating their range	LAS specification 1.4 [i.41]

Depending on the type of data that is provided by sensors, not only the value of the parameter itself is of interest but also the knowledge under which circumstances, when are where, this parameter had been sampled. This type of information, the so called meta data, depends on the characteristics of the sensor and also on the intended usage of the sensor data as part of an application. Classification schemes and formatting for several purposes are existing and Table 4 lists some examples.

Table 4: Classification schemes and formatting for sensor data

Protocols	SDO	Features	Standards
Geographic information: Observations and measurements	ISO	Conceptual schema for observations and features involved, exchange of information	ISO 19156 [i.25]
OGC® SensorML: Model and XML Encoding Standard	OGC®	Definition of a model and an XML schema for sensors	OGC® SensorML [i.37]: Model and XML Encoding Standard
OGC® Sensor Things API Part 1: Sensing	OGC®	Definition of a framework for interconnecting and monitoring IoT devices with APIs	OGC® Sensor Things API Part 1: Sensing [i.38]
OGC® Sensor Things API Part 2: Tasking Core	OGC®	Definition of a framework for parameterizing IoT devices with APIs	OGC® Sensor Things API Part 2: Tasking Core [i.39]
MPEG-V: Part 3: Sensory information	ISO	Description language and vocabulary to describe sensorial effects	ISO/IEC 23005-3 [i.45]
MPEG-V Part 2: Control information for actuators and sensors	ISO	Syntax and semantics to provide interoperability for actuators and sensors in real and in virtual worlds	ISO/IEC 23005-2 [i.56]

4.2.4 Standards and technologies for ARF Reference Point 3

This reference point supports the dialog for real-time communication and data exchange between the Scene Management (SM) functions and an external system to ensure that relevant features of such a system are available to SM subfunctions for processing and control.

Many AR applications are tailored to support detailed purposes and so for many use cases it is already known what type of external service to contact in case this is needed. If this information is not available or requirements to a service are flexible over time, standards are existing that support the detection of available services. Examples for such a detection of services on the network layer are listed in Table 5.

Table 5: Standards for the detection of external services

Protocols	SDO	Features	Standards
Session Announcement Protocol	IETF	Advertisement of sessions, possibility to communicate the relevant session setup information	IETF RFC 2974 [i.15]
Service Location Protocol	IETF	Framework for providing clients with access to information about networked services	IETF RFC 2608 [i.16]
Session Description Protocol (SDP)	IETF	Detection of available services using DNS queries	IETF RFC 6763 [i.17]

Many AR implementations possess application-specific information about which types of parameters more detailed information is needed in order to run the implementation in a smooth and useful way. In case of uncertainties, a capability exchange with the external service can be realized by the standards mentioned in Table 6.

Table 6: Standards for the capability exchange with external services

Protocols	SDO	Features	Standards
Diameter Base Protocol	IETF	Framework for authentication, authorization and accounting of IP services for network parameters	IETF RFC 6733 [i.18]
Message Queuing Telemetry Transport (MQTT)	Oasis	Network protocol for machine-to-machine communication using message brokers	MQTT [i.40]

4.2.5 Standards and technologies for ARF Reference Point 4

This RP supports the treatment of user interactions captured by input devices to steer, control or influence the behaviour of an AR scene. User interactions can happen via different input modalities. The ARF architecture differentiates 3D-gesture-based, tactile, gaze, voice, and biometric interaction. This list may be extended in the future with new interaction modalities to appear in the technological landscape. The listed categories can be subdivided further with respect to concrete approaches. For example, a tactile interaction can take place via touch-responsive surfaces, knobs, buttons, or interactive real-world objects in general. Similarly, biometric interaction can be based on heart rate, Electroencephalogram, or skin-resistance. For corresponding devices and sensors the data structures and connectivity depend on the respective vendor.

EXAMPLE 1: Some devices allow direct access to the data while others transmit the data to a cloud infrastructure, where it can be retrieved from afterwards.

Furthermore, the devices provide the data on multiple levels of fidelity.

EXAMPLE 2: Two different devices can both incorporate an infrared camera to determine a user's hand, but one provides access to the raw data of the sensor i.e. infrared images, while the other only provides access to poses of the hand and its bones. Some devices are even capable of recognizing certain interaction patterns from their captured data, e.g. complete gestures users perform with their hands.

Therefore, the relevant APIs and data formats with respect to user interactions are manyfold and strongly depend on the applied input modalities, variants, vendors, and the level of data processing taking place in an input device. They also change on a regular basis due to the technological evolution of devices, new releases and updates, or older devices being discontinued. There are only few standards focusing on harmonising the breadth of available interfaces. They are implemented by several vendors and cover different types of input. However, the vendors may wrap the standards into their own particular API, e.g. to integrate a standard into an existing framework. By doing so, a standard is used and implemented by the vendor, but not explicitly available to the users of a vendor's device, game engine, or development environment.

Table 7 lists those standards, that are supporting platform-independent user interactivity. Other platform or vendor specific solutions are not considered in the present document.

Table 7: Standards for supporting user interactivity

Protocols	SDO	Features	Standards
MPEG-V Part 5: Data formats for interaction devices	ISO	Description of a schema for data formats for interaction devices	ISO/IEC 23005-5 [i.26]
MPEG-U Part 2: Interface for advanced user interaction	ISO	Description of schemes for data formats to transmit user interactions in form of shapes, postures, gestures, touch, and composite patterns; focuses on higher level interaction data	ISO/IEC 23007-2 [i.49]
WebXR Device API: Input	W3C®	Definition of a platform-agnostic mapping between vendor specific input devices and a generic interaction framework to react on user input	WebXR Device API [i.50]
Open XR: Input Actions	Khronos®	Definition of a platform-agnostic mapping between vendor specific input devices and a generic interaction framework to react on user input	OpenXR™ specification [i.42], clause 11
Graphics Language Transmission Format	Khronos®	Propose behaviour graphs for enabling content creators to add logic and behaviours to glTF assets	ISO/IEC 12113 [i.27] - glTF™ 2.0 Interactivity Extension (KHR_interactivity)

4.2.6 Standards and technologies for ARF Reference Point 5

This RP supports the representation of the AR scene in a media-related way. The data formats for the presentation of the AR/XR scene are often implementation-specific and as such not standardized and also very often not described. However, taking into account that presentation devices as head-mounted displays can be used for different applications and also often possess built-in processing power, it is possible that this RP conveys compressed data for audio-visual media. If such capabilities exist, the technologies and standards listed in Table 2 can be used.

For initiating and modifying the presentation of rendered media, the technologies listed in Table 8 can be used.

Table 8: Protocols for Capability Exchange

Technologies	SDO	Features	Standards
Session Description Protocol (SDP)	IETF	Textual description of media-related parameters	IETF RFC 3866 [i.84]
Media Capabilities	W3C®	API description to allow browsers to make an optimal decision when picking media content	Media Capabilities [i.4]
Media Source Extensions	W3C®	API description for the dynamic construction of media streams for browsers	Media Source Extensions [i.5]
OpenXR™: Rendering	Khronos®	API description for application-specific presentation of images using different XR platforms	OpenXR™ specification [i.42], clause 10

4.2.7 Standards and technologies for ARF Reference Point 6

This RP supports the gathering of information about the rendering and playback quality for audio-visual media. Many error types are visible or audible to the user, but this type of error is not automatically signalled to the Scene Management. Metrics concentrate on numerically measurable events. Such signalling is possible, if the 3D Rendering possesses some knowledge about the characteristics of the media and has the possibility to measure discrepancies between real and expected values as e.g. data rate, frame rate, packet sizes or unexpected media profiles. Occurring discrepancies need to be categorized and it needs to be decided whether these discrepancies influence in a negative way the rendering quality. For video elements the technology listed in Table 9 can be used.

Table 9: Protocols for Playback Metrics

Technology	SDO	Features	Standards
Media Playback Quality	W3C®	Extension of the HTMLVideoElement to detect user perceived quality aspects	Media Playback Quality [i.6]

4.2.8 Standards and technologies for ARF Reference Point 7

This RP supports the description of the AR scene with all objects and their behaviour. Depending on the purpose of the application and the complexity of the scene, such descriptions can range from additional information about the surface characteristics of 3D models up to a description of a complete AR scene including the possibilities for interaction between the objects that make up the scene. Some of the formats used, are listed in Table 10. All information is stored in files, very often as ASCII descriptions, sometimes also as a combination of ASCII-based and binary information.

Table 10: Schemata for scene descriptions

Specifications	SDO / Provider	Features	Standards
OBJ File Format	Open file format, originally developed by Wavefront Technologies	Storage of geometric objects composed of lines, polygons, and free-form curves and surfaces	Wavefront OBJ File Formats [i.44]
Collada™	Khronos®	XML-based schema for the transport of 3D assets between applications	Collada™ [i.43]
Graphics Language Transmission Format	ISO	File format for three-dimensional scenes and models	glTF™ - ISO/IEC 12113 [i.27]
OGC® Augmented Reality Markup Language	OGC®	Describing virtual objects in an Augmented Reality scene with their appearances and their anchors related to the real world	ARML [i.57]
Virtual Reality Modelling Language (VRML)	ISO	File format for describing interactive 3D objects and worlds for the web	ISO/IEC 14772-1 [i.58]
ISO/IEC 19775-1: Extensible 3D (X3D), architecture and components	ISO	File format for publishing, viewing, printing and archiving interactive 3D models on the Web	ISO/IEC 19775-1 [i.55]
Open Metaverse Interoperability Group (OMI)	W3C®	Extensions of glTF™ for the management of interactive scenes	OMI [i.59]
Universal Scene Description (USD)	Open file format created by Pixar Animation Studios	File format for interchange of 3D computer graphics data	USD [i.60]
Autodesk Filmbox (FBX)	Closed file format developed by Autodesk®	3D authoring and interchange format that provides access to 3D content from most 3D vendors	FBX [i.61]
MPEG-A Part 13: Augmented Reality Application Format (ARAF)	ISO	Data format used to provide an augmented reality presentation	ARAF [i.62]
MPEG-4 Animation Framework eXtension (AFX)	ISO	General organization of synthetic models in a six-component hierarchy: geometry, modelling, physical, biomechanical, behavioural and cognitive component	AFX [i.63]
MPEG-I: Scene Description for MPEG Media	ISO	Extensions of glTF™ for including rich media and interactive content	MPEG-I [i.64]

4.2.9 Standards and technologies for ARF Reference Point 8

The recognition of an object and the determination of its location and position is an important step for initiating an AR application in a proper way. This RP supports the information about the position and the orientation of virtual or real-world objects. The used methodologies are often application-specific and not standardized. However, there are data structures specified to convey data about the pose of an object and also about additional metadata that can be used to handle such information in different environments. An overview about widespread data structures is given in Table 11.

Table 11: Data structures for the pose of objects

Specifications	SDO / Provider	Features	Standards
OpenXR™: Rendering	Khronos®	Data structure of a pose and its handling in different use cases	OpenXR™ specification, (mainly in sections 2.16 "Coordinate systems" and 6.2.3 "Standard pose identifiers") [i.42]
OGC® GeoPose	Open Geospatial Consortium	Requirements for the interoperable exchange of the location and orientation of real or virtual geometric objects	OGC® GeoPose [i.46]
MPEG-4 Part 11: Scene Description and Application Engine	ISO	Scene description for MPEG-4 environments	ISO/IEC 14496-11 [i.47]
WebXR Device API	W3C®	Web support for rendering 3D scenes to hardware, designed for presenting virtual worlds in AR or VR	WebXR [i.65]
ARKit	Apple®	Apple-only AR API with pose estimation, object detection, user and image tracking, meshing	ARKit [i.66]
ARCore	Google®	Google's augmented reality SDK offering cross-platform APIs with pose estimation, object detection, user and image tracking, persistent anchors	ARCore [i.67]
Snapdragon Spaces	Qualcomm®	Cross platform API for creating AR experiences with pose estimation, object detection, user and image tracking, persistent anchors, meshing	Snapdragon spaces [i.68]

4.2.10 Standards and technologies for ARF Reference Point 9

This RP supports the information about objects, the so called metadata, of the real world that have been recognized or identified as part of the AR scene. As such, the data flow across this RP is application-specific. However, there are recommendations for structuring the metadata in such a way that specific features or attributes of these objects are categorized to facilitate their usage as part of the application.

A non-exhaustive collection of metadata specifications is given in Table 12.

Table 12: Specifications for the structuring of metadata

Specifications	SDO	Features	Standards
ISO/IEC 11179-3: Registry metamodel and attributes	ISO	Model for metadata and basic attributes	ISO/IEC 11179-3 [i.51]
The Dublin Core Metadata Element Set	IETF	Metadata elements for resource description in a cross-disciplinary information environment	IETF RFC 5013 [i.52]
ISO 19115-1: Fundamentals for metadata for geographic information	ISO	Metadata schema for describing geographic information	ISO 19115-1 [i.53]
ISO/IEC 19139-1: XML schema for ISO 19115-1 (geographic metadata)	ISO	XML schema implementation for geographic metadata as specified in ISO 19115-1 [i.53]	ISO/TS 19139-1 [i.54]

4.2.11 Standards and technologies for ARF Reference Point 10

This RP supports the 3D information about the geometry of the real world and the model that is used to describe it. A non-exhaustive list of technologies that can be used considering 3D information is given in Table 13.

Table 13: Methods for creating and representing 3D information

Protocols	SDO	Features	Standards/Technologies
ISO/IEC 19775-1: Extensible 3D (X3D), architecture and components	ISO	Declarative representation of 3D computer graphics describing the architecture and the components of an X3D graphic	ISO/IEC 19775-1 [i.55]
ARKit	Apple	Meshing and scene objects classification, plane detection of flat surfaces	ARKit [i.66]
ARCore	Google	Plane detection of flat surfaces	ARCore [i.67]
Snapdragon Spaces	Qualcomm	Meshing and scene objects classification, plane detection of flat surfaces	Snapdragon [i.68]

4.2.12 Standards and technologies for ARF Reference Point 11

This RP supports information regarding the relocalization of an element of the real environment to handle within the AR application. Due to the fact that AR applications augment reality with additional information and the provisioning of object-related processes, a proper relocalization is crucial for the quality of the user experience. Depending on the usage scenario of the application, different technologies are used for the relocalization process and this is also the reason why there are only proprietary solutions available.

For visual relocalization the analysis of still images or videos is widespread with the goal to extract characteristic features from these sources and to compare them with existing material for matching the objects of interest. For this task, very often neural networks are used to detect characteristic features in the source material and to compare them to already existing ones in a repository. This comparison can be facilitated by additional information as e.g. the geo position of the recording camera, its focal length, its picture formats and other calibration parameters.

4.2.13 Standards and technologies for ARF Reference Point 12

This RP supports information about the recognition and identification of objects in the real world as part of the global world representation. For audio-visual content there are descriptors available (see Table 14) that can support the analysis of content in the World Analysis and these descriptors can be sent to the World Storage to speed up search in large repositories.

Table 14: Descriptors for audio-visual content

Specifications	SDO	Features	Standards
MPEG-7 Part 13: Compact descriptors for visual search	ISO	Descriptors for the analysis of still images	ISO/IEC 15938-13 [i.69]
MPEG-7 Part 15: Compact descriptors for video analysis	ISO	Descriptors for the analysis of objects in video streams	ISO/IEC 15938-15 [i.70]
MPEG-7 Part 4: Descriptors for audio analysis	ISO	Descriptors for the analysis of audio signals	ISO/IEC 15938-4 [i.71]

Depending on the technologies used, the recognition of objects in the real world can also be supported by the file format that is used to represent these objects. For a proper presentation of 3D objects, information about a coordinate system needs to be added to each element (e.g. a vertex in a polygon model or a data point in a point cloud) Many of the file formats are purely text-based and therefore easy to inspect and to edit. Their structure however is proprietary and adapted to the necessary features of the application.

4.2.14 Standards and technologies for ARF Reference Point 13

This RP supports the provisioning of objects of an AR scene and updates and information about their relationship for use by the Scene Management. All the schemata that are listed in Table 10 can also be used to update the scene description depending on the actual user journey but also on technical requirements as e.g. available bandwidth, processing power of the device or regulations on storage capacities.

The connection between the client that holds the scene description to be updated and the server that holds new elements bases on the use of the HTTP protocol (see IETF RFC 9113 [i.74]). In combination with this protocol, mechanisms like WebSockets (see IETF RFC 6455 [i.75]) or Remote Procedure Calls (see IETF RFC 5531 [i.76]) are used to establish the communication, to control data transfer and to close the connection again. For transmission purposes, the serialization of the data representing an object is done using JSON.

4.2.15 Standards and technologies for ARF Reference Point 14

This RP supports information captured at runtime during an AR session regarding the behaviour and the quality of the AR experience of the presented AR scene, the user's interactions, or the status of the system itself for use by the Asset Preparation. An overall quality estimation of an AR scene is very difficult to do, due to the unknown mixture of real-world and virtual-world elements and due to the fact that there is no fixed viewing angle during the presentation of the scene. Runtime environments are able to monitor the quality of selected events or elements of the AR scene and can either directly try to react to it or send some status information to other components. Such type of quality monitoring needs to be installed as part of the scene description during its development. The signalling format of quality parameters depends on the used development environment.

Due to the manifold possibilities regarding the design of an AR scene, there is no standardized set of events and accompanying parameters to be used. The designer of an AR scene needs to decide what type of events needs to be monitored at what level of detail. The monitoring of events like the availability of the network connection, the interruption of the presentation of a session or the minimum acceptable frame rate of images caught by integrated cameras is often established as part of the scene design.

4.2.16 Standards and technologies for ARF Reference Points 15 and 19

These reference points support the provisioning of the 3D representation of objects that are reconstructured from observations of the real world or based on pre-existing models to be used as part of the scene design. This 3D representation can be added to an AR scene in the AR Authoring function. Some of the file formats that can be used for this task are listed in Table 15. Some technologies that can provide the world representation for AR applications are listed in Table 16.

Table 15: File formats for 3D objects

Protocols	SDO	Features	Standards
OBJ File Format	Open file format	Storage of geometric objects composed of lines, polygons, and free-form curves and surfaces.	Wavefront OBJ File Formats [i.44]
Autodesk Filmbox (FBX)	Closed file format developed by Autodesk®	3D authoring and interchange format that provides access to 3D content from most 3D vendors. SDKs are available to convert such files into other formats	FBX [i.61]
ISO/IEC 19775-1: Extensible 3D (X3D), architecture and components	ISO	3D graphics file format based on XML to store representations of 3D objects and scenes	ISO/IEC 19775-1 [i.55]
Graphics Language Transmission Format	ISO	File format for three-dimensional scenes and models	glTF™ - ISO/IEC 12113 [i.27]
Universal Scene Description (USD)	Open file format created by Pixar Animation Studios	File format for interchange of 3D computer graphics data	USD [i.60]
3D Tiles	Cesium	Open standard for massive, heterogeneous 3D geospatial datasets such as point clouds, buildings, and photogrammetry based on glTF™	3D Tiles [i.82]

Table 16: Technologies for accessing the World Representation

SDO/Publisher	Technology	Internal reference
Apple®	RoomPlan	See clause 5.17.2, RoomPlan
Apple®	ARKit	See clause 5.17.1, ARKit
Google®	ARCore	See clause 5.18.1, ARCore
Cesium Platform®	Cesium	See clause 5.20.1, Cesium platform
Qualcomm®	Snapdragon spaces	See clause 5.19.1, Snapdragon Spaces

4.2.17 Standards and technologies for ARF Reference Point 16

This RP supports the provisioning of several anchors in order to attach virtual content to them. The position of these anchors is defined by coordinate systems. The provisioning of coordinate systems and the information about the elements in the real world to which these coordinate systems are related is done by this reference point. There are thousands of reference systems in use but most of them based on common strategies that were defined by standards that are listed in Table 17.

Table 17: Definition of reference systems

Protocols	SDO	Features	Standards
Geographic information: Referencing by coordinates	ISO	Conceptual schema for referencing by coordinates	ISO 19111 [i.77]
OGC® Referencing by coordinates Corrigendum	OGC®	Corrigendum and extension to ISO 19111 [i.77] considering reference coordinate systems	OGC® Referencing by coordinates [i.78]
OGC® API - Coordinate Reference Systems	OGC®	Usage of coordinate reference identifiers	OGC® API Coordinate Reference Systems [i.79]
Geographic information: Referencing by geographic identifiers	ISO	Conceptual schema for referencing by geographic identifiers	ISO 19112 [i.80]

4.2.18 Standards and technologies for ARF Reference Point 17

This RP supports the provisioning of an object as representation of an element of the real world for relocalization (e.g. 2D markers or 3D models), recognition or tracking purposes with the following functionalities: Provisioning of new objects representing a physical element of the real world for an AR scene, provisioning of both pose and scale of objects in relation to a coordinate system of the real world, the updating of properties of objects used for the representation of the real world and the deletion of objects from the representation of the real world.

The data formats that are used to convey such information depend on the technologies used to design the AR scene. Technologies and standards for such a type of representation are already listed in the present document. References to the appropriate tables and clauses are as follows:

- For positioning technologies see Table 3, for the pose of objects see Table 11 and for reference systems see Table 17.
- The representation of an element as a 3D model or as a planar image can de done with the standards listed in Table 18.

Table 18: Representation of elements

Type of representation	SDO / Publisher	Technology	Internal reference
3D model	Wavefront	OBJ File Format	See clause 5.14.1, OBJ File Format
3D model	ISO	Graphics Language Transmission Format glTF	See clause 5.3.9, Graphics Language Transmission Format
3D model	OGC®	Augmented Reality Markup Language ARML	See clause 5.10.5, OGC® Augmented Reality Markup Language
3D model	ISO	Virtual Reality Markup Language VRML	See clause 5.3.18, Virtual Reality Modelling Language (VRML)
3D model	Pixar	Universal Scene Description USD [i.60]	See clause 5.15.1, Universal Scene Description (USD)

Type of representation	SDO / Publisher	Technology	Internal reference
Planar image	ITU	JPEG	See clause 5.4.3, Digital Image Compression JPEG
Planar image	W3C®	Scalable Vector Graphics SVG	See clause 5.1.2, Scalable Vector Graphics
Planar image	ISO/IEC	Portable Network Graphics PNG	See clause 5.3.11, Portable Network Graphics (PNG)

4.2.19 Standards and technologies for ARF Reference Point 18

This RP supports the provisioning of all assets required to author an AR scene consisting of scene description, the provisioning of synthetic and/or audio-visual elements and their behaviour as part of the scene.

The data formats that are used to convey such information depend on the technologies used to design the AR scene. Technologies and standards for such a provisioning are already described in the present document. Table 19 gives the references to document-internal descriptions considering scene descriptions and their management.

Table 19: Technologies for scene descriptions

SDO/Publisher	Technology	Internal reference
Pixar	Universal Scene Description USD [i.60]	See clause 5.15.1, Universal Scene Description (USD)
Autodesk	Autodesk Filmbox FBX [i.61]	See clause 5.16.1, Autodesk Filmbox (FBX)
Apple	ARKit	See clause 5.17.1, ARKit
Google	ARCore	See clause 5.18.1, ARCore
MPEG-4 Part 11	Binary formats for scenes	See clause 5.3.10, MPEG-4 Part 11: Scene Description and Application Engine
MPEG-I	Scene description for MPEG media	See clause 5.3.21, MPEG-I: Scene Description for MPEG Media
MPEG-A	Augmented Reality Application Format	See clause 5.3.19, MPEG-A Part 13: Augmented Reality Application Format (ARAF)
OGC®	Augmented Reality Markup Language for scene integration	See clause 5.10.5, OGC® Augmented Reality Markup Language

Table 20 gives the references to document-internal descriptions considering the representation of audio-visual respectively synthetical elements as part of the AR scene.

Table 20: Technologies for audio-visual and synthetic scene elements

SDO/Publisher	Technology	Internal reference
ISO	Virtual Reality Markup Language VRML	See clause 5.3.18, Virtual Reality Modelling Language (VRML)
ISO	Extensible 3D	See clause 5.3.16, ISO/IEC 19775-1: Extensible 3D (X3D), architecture and components
Wavefront	OBJ File Format	See clause 5.14.1, OBJ File Format
W3C®	Scalable Vector Graphics SVG	See clause 5.1.2, Scalable Vector Graphics
ISO/IEC	Portable Network Graphics PNG	See clause 5.3.11, Portable Network Graphics (PNG)
ITU	JPEG	See clause 5.4.3, Digital Image Compression JPEG
ISO/IEC	V3C/V-PCC for Point Clouds	See clause 5.3.3, Visual volumetric video-based coding; Video-based point cloud compression (V3C/V-PCC)
ISO/IEC	V3C/MIV MPEG Immersive Video	Seeclause 5.3.4, Visual volumetric video-based coding; MPEG Immersive Video (V3C/MIV)
ITU	HEVC Video Coding	See clause 5.4.1, High Efficiency Video Coding (HEVC)
AOMedia	AV1 Video Coding	See clause 5.5.1, AOMedia Video 1 (AV1)
ISO/IEC	AAC Advanced Audio Coding	See clause 5.3.1, Advanced Audio Coding (AAC)
ISO/IEC	MPEG-H 3D Audio	See clause 5.3.2, MPEG-H 3D Audio

5 Standards Ordered by Standard Development Organizations

5.1 W3C®

5.1.1 WebRTC: Real Time Communication between Browsers

This recommendation [i.2] describes a set of APIs to allow media and generic application data to be sent to and received from another browser. The document describes mechanisms allowing direct audio and video communication to work within web pages without the need to install plugins or native applications. Underlying technical enablers are HTML and JavaScript.

Internet link to the document: WebRTC

5.1.2 Scalable Vector Graphics

This recommendation [i.3] specifies the features and the syntax for defining two-dimensional graphics by a format that uses the markup language XML. The specification supports scaling, interactivity and animation. As the representation of graphics in XML is done text-based, the format is lossless. However, the definition of graphics in an XML file contains many repeating elements. This redundancy can be eliminated by using lossless compression techniques for text-based files as e.g. zip.

Internet link to the document: SVG 1.1

5.1.3 Media Capabilities

This specification [i.4] provides a set of APIs to allow a browser to make an optimal decision when selecting audio or video media based on information such as the codecs, profile, resolution, bitrates, etc. The APIs expose information about the decoding and encoding capabilities for a given format but also output capabilities to find the best match based on the characteristics of the display. Real time feedback is available to support adaptive streaming to modify the quality of the content based on actual user perceived quality.

Internet link to the specification: Media Capabilities

5.1.4 Media Source Extensions

This specification [i.5] allows the creation of dynamical media streaming for audio and video elements to be presented by HTML5 elements within a browser. It defines a MediaSource object that can serve as a source of media data for an HTML media element. These MediaSource objects make use of one or more buffer objects, the so called source buffers. Applications append data segments to these buffers, and can adapt the quality of appended data based on system performance and other factors.

Internet link to the specification: Media Source Extensions

5.1.5 Media Playback Quality

This specification [i.6] describes extensions to the HTMLVideoElement to gather information about user perceived quality aspects.

Internet link to the specification: Media Playback Quality

5.1.6 Web of Things Thing Description

This standard [i.7] defines the characteristics of a W3C® Thing. The description of a Thing consists of four main components: textual metadata about the Thing itself, a set of interaction descriptions that indicate how the Thing can be used, schemas for the data exchanged with the Thing for machine-understandability, and web links to express any formal or informal relation to other Things or documents on the web. The description uses metadata for different protocol bindings that carry an implementation of a thing description. The serialization of a description instance is done using JSON.

Internet link to the standard: Web of Things Thing (WoT) Description

5.1.7 WebXR Device API: Input

This W3C® candidate recommendation [i.50] specifies several interfaces to hardware that enables VR and AR applications offering to the user an immersive computing platform with new and extended functionalities. In such an environment the ability to interact directly with immersive hardware is critical to ensuring that the applications can operate in a precise and also user-friendly way. The WebXR Device API provides the interfaces necessary to enable developers to build compelling, comfortable, and safe immersive applications on the web across a wide variety of hardware form factors. Section 10 of this recommendation specifies those parameters that allow the user to perform targeted actions in the space provided by an application.

Internet link to section 10 of the specification: WebXR Device API; Input

5.1.8 Open Metaverse Interoperability Group (OMI)

The Open Metaverse Interoperability Group [i.59] is focused on bridging virtual worlds by designing and promoting protocols for identity, social graphs, inventory, and more. It is associated with the Metaverse Interoperability community group at W3C®. The group develops and documents extensions for the glTF™ format for the management of interactive scenes: physics, virtual agents, spawning, audio, etc.

Internet link the specifications of the glTF™ extensions: OMI Github

5.1.9 WebXR Device API

WebXR [i.65] is a group of standards which are used together to support rendering 3D scenes to hardware designed for presenting virtual worlds (virtual reality, or VR), or for adding graphical imagery to the real world (augmented reality, or AR). WebXR-compatible devices include fully-immersive 3D headsets with motion and orientation tracking, eyeglasses which overlay graphics atop the real-world scene passing through the frames, and handheld mobile phones which augment reality by capturing the world with a camera and augment that scene with computer-generated imagery. The WebXR Device API provides the following key capabilities: find compatible VR or AR output devices, render a 3D scene to the device at an appropriate frame rate, mirror the output to a 2D display and create vectors representing the movements of input controls.

Internet link the specification: WebXR

5.2 IETF

5.2.1 Transmission Control Protocol (TCP)

TCP provides the delivery of a stream of packetized data between applications running on hosts and communicating via an IP network. The protocol (IETF RFC 9293 [i.8]) is connection-oriented and as such a connection between the hosts is established before data is sent. The Transmission Control Protocol accepts data from a data stream, divides it into chunks, and adds a TCP header to build a so called TCP segment. By using error detection schemes, the protocol supports a reliable transmission of data in an ordered way. In case of failures, a retransmission is done. These mechanisms improve the reliability of the connection but also adds latency.

Internet link to the standard: IETF RFC 9293

5.2.2 User Datagram Protocol (UDP)

UDP provides the delivery of a stream of packetized data between applications running on hosts and communicating via an IP network. The protocol (IETF RFC 768 [i.9]) is connectionless and as such each transmitted data packet is individually addressed and routed across the network. It does not require the set-up of communication channels before transmission starts. As for TCP, the protocols accept data from a stream, divides it into chunks, and adds a UDP header to it to build a so called datagram. The protocol supports the creation of check sums for each datagram for integrity reasons but there are no mechanisms available for error correction. The protocol is often used for applications where no error checking or correction is needed or is done as part of the application. The lack of time-consuming checking mechanisms makes this protocol a preferred one for the support of real-time applications.

Internet link to the standard: IETF RFC 768

5.2.3 Internet Protocol (IP)

The Internet Protocol is responsible for transmitting packetized data, so called datagrams, between hosts across network boundaries. The source and the destination host are described by IP addresses and the protocol makes use of these addresses and other additional data to route and to deliver datagrams to the destination. The assignment of IP addresses is worldwide done in an unambiguous way and these addresses are publicly available for users of the Internet. With the assignment of IP addresses to a multitude of devices over the last decades, the address range needed to be extended. The primary protocol version (IPv4) (IETF RFC 791 [i.10]) works with 32 bits per IP address, the extended version (IPv6) (IETF RFC 8200 [i.11]) can make use of up to 128 bits per IP address.

Internet links to both standards: IETF RFC 791 (IPv4), IETF RFC 8200 (Ipv6)

5.2.4 Real-time Transmission Protocol (RTP)

With the help of the RTP protocol (IETF RFC 3550 [i.12]), audio and video streams are delivered over IP networks. It typically runs over the UDP protocol and is designed for a real-time, end-to-end transfer of streaming media. It provides opportunities for jitter reduction and for the detection of packet loss and delivery of data packets in a different order from which they were sent. To support these features, the protocol includes timestamps for synchronization, sequence numbers and information about the encoded format of the streaming data.

Internet link to the standard: IETF RFC 3550

5.2.5 RTP Control Protocol (RTCP)

The RTP Control Protocol (IETF RFC 3550 [i.12]) is used in combination with the RTP protocol. It provides control information for session management and statistics on quality aspects of the media distribution during a session and transmits this data to the session media source and other session participants. By using this protocol, all participants of a session can be reached. Information that is distributed by the sender is synchronized to the media distribution by using the same timestamp for both protocols RTCP and RTP.

5.2.6 Real-time Streaming Protocol (RTSP)

The RTSP protocol (IETF RFC 7826 [i.13]) is designed for multiplexing and packetizing multimedia transport streams over a suitable transport protocol. The protocol operates between clients and servers, but it also supports the use of proxies placed between clients and servers. Clients can request information about streaming media from servers by asking for a description of the media at the location of the media server or to use media description that is provided externally.

Internet link to the standard: IETF RFC 7826

5.2.7 Session Initiation Protocol (SIP)

The SIP protocol (IETF RFC 3261 [i.14]) is used for creating, maintaining and terminating communication sessions with one or more participant. The protocol defines the specific format of messages exchanged and the sequence of communications for cooperation of the participants. It is text-based and most commonly, the negotiation and the media setup is done with the Session Description Protocol that is carried as a payload of SIP messages. The protocol is designed to be independent of the underlying transport protocol, mainly UDP and TCP are used.

Internet link to the standard: IETF RFC 3261

5.2.8 Session Announcement Protocol

IETF RFC 2974 [i.15] specifies the advertisement of sessions, and the possibilities to communicate the relevant session setup information to prospective participants by using well-known multicast addresses. The packet format for this protocol is specified; the information is periodically distributed by a session directory. The protocol does not contain any rendezvous mechanism, the announcer of a session is not aware of any listeners.

Internet link to the standard: IETF RFC 2974

5.2.9 Service Location Protocol

IETF RFC 2608 [i.16] specifies a flexible and scalable framework for providing hosts with access to information about the existence, location, and configuration of networked services. Normally, users have had to find services by knowing the name of a network host which is an alias for a network address. This protocol eliminates the need for a user to know the name of a network host supporting a service. Instead, the user supplies the desired type of service and a set of attributes which describe the service. Based on that description, the Service Location Protocol resolves the network address of the service for the user.

Internet link to the standard: IETF RFC 2608

5.2.10 DNS-Based Service Discovery

IETF RFC 6763 [i.17] specifies how DNS resource records are named and structured to facilitate service discovery. Given a type of service that a client is looking for, and a domain in which the client is looking for that service, the mechanism described in this standard allows users to discover a list of named instances of that desired service, using standard DNS queries, and then select, from that list, the particular instance they are interested in.

Internet link to the standard: IETF RFC 6763

5.2.11 Diameter Base Protocol

IETF RFC 6733 [i.18] provides a framework for authentication, authorization and accounting of applications such as network access or IP mobility. It specifies the message format, transport, error reporting, accounting, and security services used by all Diameter applications. One of the supported features is the capability exchange between clients related to network parameters. All data exchange between clients happens via Attribute-Value Pairs (AVPs).

Internet link to the standard: IETF RFC 6733

5.2.12 Session Description Protocol (SDP)

IETF RFC 8866 [i.19] specifies the syntax and the elements of a protocol that is intended for describing multimedia sessions by supporting session announcement, session invitation, and other forms of multimedia session initiation. Parameters are also available for media descriptions (e.g. type of media, resolution, repetition rate). All media-related clients of a session offer their technical capabilities and give priorities to them if more than one format is available. By a handshake procedure all involved clients can agree on the formats to be used.

Internet link to the standard: IETF RFC 8866

5.2.13 The Dublin Core Metadata Element Set

IETF RFC 5013 [i.52] specifies a vocabulary of fifteen properties for use in resource description. The resources described using this set may be digital resources (e.g. video, images, web pages) as well as physical resources. Each Dublin Core element is optional and may be repeated. There is no prescribed order in the Dublin Core metadata element set for presenting or using these elements.

Internet link to the standard: IETF RFC 5013

5.3 ISO

5.3.1 Advanced Audio Coding (AAC)

The AAC codec [i.20] achieves data compression by discarding signal components that are perceptually irrelevant and by eliminating redundancies in the coded signal. Depending on the complexity of the signal to be encoded, implementers can select between several profiles. Sampling rates up to 96 kHz can be supported for up to 48 audio channels. Different compression schemes can be used for different frequency ranges resulting in arbitrary bit rates, typically in the range between a few kilobits per second up to a few hundreds of kilobits per second and channel.

Internet link to the standard: ISO/IEC 14496-3:2019

5.3.2 MPEG-H 3D Audio

This codec (ISO/IEC 23008-3 [i.21]) supports the representation of audio as audio channels, as audio objects and as full-sphere surround sound. This plurality is achieved by working with audio representation as a sound field in an intermediate data format that is rendered in a following processing step to the speaker setup. The codec supports up to 64 loudspeaker channels. The availability of processing audio objects allows interactivity and personalization of an audio scene.

Internet link to the standard: ISO/IEC 23008-3:2022

5.3.3 Visual volumetric video-based coding; Video-based point cloud compression (V3C/V-PCC)

This codec (ISO/IEC 23090-5 [i.22]) supports the representation and the compression of volumetric video consisting of volumetric frames where each frame is a three-dimensional grid of voxels. Each voxel represents a characteristic value within this grid and can be accompanied by additional metadata. For this encoding scheme the three-dimensional grid is represented as a point cloud. Each point is associated with its geometrical position, together with associated attribute information such as colour, reflectance, and transparency. The compression of the point cloud is done with the support of traditional two-dimensional video coding techniques. To achieve this, each volumetric frame is converted to a number of two-dimensional planes by using virtual cameras. Each of these planes contributes to the two-dimensional frame representation. Information about their position, their dimensions and the position of the virtual cameras can also be used to used to restore the volumetric frames again.

Internet link to the standard: ISO/IEC 23090-5:2023

5.3.4 Visual volumetric video-based coding; MPEG Immersive Video (V3C/MIV)

This codec (ISO/IEC 23090-12 [i.73]) supports the compression of multi-view, multi-view plus depth and multi-plane image source data. The technique was developed to support compression of immersive video content, in which a real or virtual 3D scene is captured by multiple real or virtual cameras. The standard allows the storage and distribution of immersive video content over different types of networks, for playback with 6 Degrees of Freedom (6DoF) of view position and orientation. Using HEVC Level 5.2, this compression scheme can achieve bit rates around 15 Mb/s to 30 Mb/s. For presentation equirectangular, perspective, or orthographic projection is supported.

Internet link to the standard: ISO/IEC 23090-12:2023

5.3.5 Radio Frequency Identification (RFID)

This technology (ISO/IEC 18000-3 [i.23] and ISO/IEC 15961-1 [i.24]) makes use of radio signals to identify and track tags that are attached to objects. Especially in static environments this technology is often used to identify objects the position of which is known. Tags can be active or passive. Passive tags have no internal power source. They work by inductive or capacitive coupling systems between the tag and the receiving device to transmit their signal. Active tags possess their own power supply and broadcast their signal. The reader only relies on the transmitted information of the tag, an evaluation of received signal strength or a processing of the receiving angle does not take place. Passive tags can be cheaply produced and the necessity for a narrow coupling for data transmission supports the accuracy for locating the object that carries the tag.

Internet link to the standard for parameters for air interface: ISO/IEC 18000-3:2010

Internet link to the standard for RFID data protocol: ISO/IEC 15961-1:2021

5.3.6 Geographic information: Observations and measurements

ISO 19156 [i.25] defines a conceptual schema for observations, and for features involved in sampling when making observations, including relationships between sampling features. The document also provides a model for the exchange of information describing observation acts and their results. For the presentation of the schema UML is used. The classification of features is primarily done using topological aspects.

Internet link to the standard: ISO 19156:2023

5.3.7 MPEG-V: Part 3: Sensory information

ISO/IEC 23005-3 [i.45] specifies syntax and semantics of tools describing sensory information to enrich audio-visual contents. A Sensory Effect Description Language as an XML schema-based language enables the description of a basic structure of sensory information, and a Sensory Effect Vocabulary is specified as an XML representation for describing sensorial effects such as light, wind, fog, vibration, etc. that trigger human senses.

Internet link to the standard: ISO/IEC 23005-3:2019

5.3.8 MPEG-V Part 5: Data formats for interaction devices

ISO/IEC 23005-5 [i.26] specifies syntax and semantics of the data formats for interaction devices by providing a standardized format for interfacing actuators and sensors by defining an XML schema-based language. It provides a basic structure with common information for communication with various actuators and sensors in consistency. A Device Command Vocabulary is defined to provide a format for commanding individual actuator, and a Sensed Information Vocabulary is defined to provide a format for holding information from individual sensors.

Internet link to the standard: ISO/IEC 23005-5:2019

5.3.9 Graphics Language Transmission Format

This file format, originally developed by the Khronos® Group, specifies a transmission format (ISO/IEC 12113 [i.27]) for three-dimensional scenes and objects. It can be represented as a binary or an ASCII/JSON file. The elements of this format support the geometry of three-dimensional objects and scenes, their appearance and animation as well as the creation of a scene graph hierarchy. The format allows the reference of external binary and texture resources. Open-source validators are available on the Internet to check the compliance of a file with the elements of this format. The format now includes a dedicated extension (KHR_Interactivity [i.85]) for managing interactivity through behaviour graphs.

Internet link to the standard: ISO/IEC 12113:2022

Internet link to the specifications of the interactivity extension: KHR_Interactivity

5.3.10 MPEG-4 Part 11: Scene Description and Application Engine

ISO/IEC 14496-11 [i.47] addresses the coding of audio-visual objects of various types: natural video and audio objects as well as textures, text, two- and three-dimensional graphics, and also synthetic music and sound effects. Besides the transmission of these objects to the terminal, additional information is needed in order to combine this audio-visual data at the terminal and construct and present to the end-user a meaningful multimedia scene. This information, called scene description, determines the placement of audio-visual objects in space and time, contains information about the interaction possibilities of the several objects and is transmitted together with the coded objects. The representation of the scene description is done in a format called "Binary Formats for Scenes" (BIFS).

Internet link to the standard: ISO/IEC 14496-11:2015

5.3.11 Portable Network Graphics (PNG)

ISO/IEC 15948 [i.48] specifies a raster-graphics file format that supports lossless data compression. PNG supports RGB-based images, and grayscale images (with or without an alpha channel for transparency). The image is represented as a series of chunks, each of which conveys certain information about the image. These chunks are characterized as critical or ancillary, allowing a program encountering an ancillary chunk that it does not understand to ignore it.

Internet link to the standard: ISO/IEC 15948:2004

5.3.12 MPEG-U Part 2: Interface for advanced user interaction

ISO/IEC 23007-2 [i.49] is part of a suite of documents that facilitates the exchange, the display, the control and the communication of components of a user interface with other entities with the aim to allow users a homogeneous, unified experience when interacting with devices. Part 2 specifies Advanced User Interaction interfaces (AUI) to support various advanced user interaction devices. The AUI interface is laid out as a bridge between scene descriptions and system resources. By a recognition process, a set of physical information can be converted to a pattern with semantics which is more useful to a scene description (e.g. recognizing a pattern of as a circle). Therefore, Part 2 provides a set of data formats which defines geometric patterns, symbolic patterns, touch patterns, posture patterns and their composite patterns.

Internet link to the standard: ISO/IEC 23007-2:2012

5.3.13 ISO/IEC 11179-3: Registry metamodel and attributes

ISO/IEC 11179-3 [i.51] specifies the structure of information suitable to be stored in a registry in the form of a conceptual data model. The constructs are semantic ones, using wider or more narrow relations between concepts and taking into account that a concept can be represented in several ways. The different possible representations of a data concept are then described with the use of one or more data elements. The standard does not describe how data will be actually stored.

Internet link to the standard: ISO/IEC 11179-3:2023

5.3.14 ISO 19115-1: Fundamentals for metadata for geographic information

ISO 19115-1 [i.53] defines the schema for describing geographic information and services by means of metadata. It contains information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services. The specification establishes a common set of metadata terminology, definitions and extension procedures.

Internet link to the standard: ISO 19115-1:2014

5.3.15 ISO/IEC 19139-1: XML schema for ISO 19115-1 (geographic metadata)

ISO/TS 19139-1 [i.54] defines XML based encoding rules for conceptual schemas specifying types that describe geographic resources. The encoding rules support the UML profile as commonly used in the UML models.

Internet link to the specification: ISO/TS 19139-1:2019

5.3.16 ISO/IEC 19775-1: Extensible 3D (X3D), architecture and components

ISO/IEC 19775-1 [i.55] describes the architecture and the base components of 3D computer graphics and defines several profiles for declaratively representing them. X3D includes multiple graphics file formats, programming-language API definitions, and run-time specifications for both delivery and integration of interactive network-capable 3D data. X3D improves upon VRML with new features, advanced application programmer interfaces, additional data encoding formats, and a componentized architecture that allows for a modular approach to supporting the standard. X3D supports variety of encoding formats and includes a rich set of features to support a wide variety of use cases including 3D graphics, animation, spatial audio, scripting, networking, programmable shaders, particles, etc. The semantics of X3D describe an abstract functional behaviour of time-based, interactive 3D, multimedia information. The document does not define any devices or implementation-specific concepts.

Internet link to the specification: ISO 19775-1:2023

5.3.17 MPEG-V Part 2: Control information for actuators and sensors

ISO/IEC 23005-2 [i.56] specifies the syntax and semantics of tools required to provide interoperability for actuators and sensors in real as well as virtual worlds. The document specifies a Control Information Description Language as an XML schema-based language which enables the description of the basic structure of control information. As XML representations it also includes a Device Capability Description Vocabulary for describing capabilities of actuators, a Sensor Capability Description Vocabulary for describing capabilities of sensors, Sensory Effect Preference Vocabulary for describing preferences of an individual user on specific sensorial effects and a Sensor Adaptation Preference Vocabulary for describing preferences of an individual user on a sensor.

Internet link to the specification: ISO/IEC 23005-2:2018

5.3.18 Virtual Reality Modelling Language (VRML)

ISO/IEC 14772-1 [i.58] defines a file format for describing interactive 3D objects and worlds. It is specified by the Web3D Consortium and has been released as the ISO/IEC 14772-1 [i.58] International Standard. VRML is designed to be used on the Internet, intranets, and local client systems. VRML is also intended to be a universal interchange format for integrated 3D graphics and multimedia. VRML is a text file format that can handle simple and complex shapes, text, images, animations, lighting, sounds, hyperlinks, as well as their spatial arrangement, texture, colour, material, etc. A dedicated script node allows the addition of program code, written in Java or ECMAScript.

Internet link to the VRML specification: ISO/IEC 14772-1:1997

5.3.19 MPEG-A Part 13: Augmented Reality Application Format (ARAF)

ISO/IEC 23000-13 [i.62] focuses on the data format used to provide an augmented reality presentation. ARAF specifies scene description elements for representing AR content, mechanisms to connect to local and remote sensors and actuators, mechanisms to integrate compressed media (image, audio, video, graphics), and mechanisms to connect to remote resources such as maps and compressed media. ARAF was developed by MPEG, the same technical committee that created mp3 for audio, AVC and HEVC for video. MPEG already provided data type representations for all kinds of media, from static image, video, audio to 3D graphics and complex dynamic scenes. Additionally, MPEG developed a set of standards related to sensors and actuators. By bringing these two components together into an application format called ARAF, MPEG enables interoperability when used to build AR applications and services.

Internet link to the standard: ISO/IEC 23000-13:2017

5.3.20 MPEG-4 Animation Framework eXtension (AFX)

The MPEG-4 Animation Framework eXtension (AFX) (ISO/IEC 14496-16 [i.63]) proposes a general organization of synthetic models in a six-component hierarchy: geometry, modelling, physical, biomechanical, behavioural and cognitive component. This conceptual organization facilitates the understanding of models and their relationships for interactive multi-media contents including (but not limited to) computer games and animation. Using this organization, AFX extends MPEG-4 Binary Format for Scenes (BIFS) features with new models for each component, dedicated bitstreams for even more compact representations, and a backchannel for view-dependent streaming of large data sets. While other frameworks provide geometry models described solely by their surface, AFX also defines solid geometry.

Internet link to the standard: ISO/IEC 14496-16:2011

5.3.21 MPEG-I: Scene Description for MPEG Media

MPEG-I part 14 scene description (ISO/IEC 23090-14 [i.64]) is a standard for the creation of augmented and virtual reality applications. It aims at creating interactive experiences with rich media such point clouds, spatial audio, immersive audio, etc. It proposes new formats and compression, also glTF™ extensions to support and access external medias (video and audio), to perform AR anchoring to detected trackables and to create interactive scenes including haptic behaviours and avatars.

Internet link to the standard: ISO/IEC 23090-14:2023

5.3.22 MPEG-7 Part 13: Compact descriptors for visual search

ISO/IEC 15938-13 [i.69] specifies descriptors for the characterization of still images. The specification is designed to enable efficient and interoperable visual search applications, allowing visual content matching in images. A set of local features are extracted from the query image and compressed into a single compact descriptor on the client side to be sent to the server for efficient search in repositories. A global descriptor is created from aggregated local descriptors of an image, allowing an extremely fast match with similar global descriptors. The standard supports different sizes of compact descriptor bit-streams, spanning from a maximum of 16 KB per image, down to 512 Bytes for extremely constrained bandwidth scenarios.

Internet link to the standard: ISO/IEC 15938-13:2015

5.3.23 MPEG-7 Part 15: Compact descriptors for video analysis

ISO/IEC 15938-15 [i.70] specifies a video description tool designed to enable efficient and interoperable visual search applications, allowing visual content matching in videos and images. This addresses matching of views of objects, landmarks and scenes, including partial temporal overlap in the video, while being robust to partial occlusions as well as changes in viewpoint, camera parameters, lighting conditions. The descriptor consists of three components, the local and the global descriptor as specified in MPEG-7 ISO/IEC 15938-13 [i.69] and a third component, the deep feature descriptor. This is extracted using a convolutional neural network, and applying a procedure called nested invariance pooling to improve its robustness to geometric transformations. The resulting descriptor is then binarized for efficient storage and matching. For compression along the temporal dimension, lossless compression is applied to the global and deep feature descriptors, while lossless or lossy compression is applied to the local feature descriptors.

Internet link to the standard: ISO/IEC 15938-15:2019

5.3.24 MPEG-7 Part 4: Descriptors for audio analysis

ISO/IEC 15938-4 [i.71] specifies parameters for describing audio content. The description is done with so called low level descriptors and the extraction of features is based on the audio signal itself. There is no semantic analysis done. With the help of seventeen temporal and spectral parameters, it is possible to search and filter audio content in regard to for e.g. spectrum, harmony, timbre and melody. Depending on the type of parameter, the description can be done as a scalar or as a vector. A sampling period of 10 ms is recommended, integer multiples of this period are also possible.

Internet link to the standard: ISO/IEC 15938-4:2002

5.3.25 Visual volumetric video-based coding; Geometry-based point cloud compression (V3C/G-PCC)

This codec (ISO/IEC 23090-9 [i.72]) supports the representation and the compression of volumetric video where three-dimensional content is directly encoded in a three-dimensional space. The coding scheme makes use of coding structures that describe point locations in three-dimensional space. The geometry and the attributes of a point cloud are encoded separately where point cloud positions are encoded first. Each point is represented by a voxel on a regular grid in three-dimensional space. For an efficient encoding of the geometry two mechanisms are described, either by a repetitive dividing of a cube into sub-cubes or by approximating the surface by a series of triangles. For the encoding of the attributes, three mechanisms are described:

- a Region-Adaptive Hierarchical Transform (RAHT);

- a Predicting Transform that exploits the Euclidian distance in the representation of attributes in a specified volume; and
- a Lifting Transform that introduces weighting factors.

Internet link to the standard: ISO/IEC 23090-9:2023

5.3.26 Geographic information: Referencing by coordinates

ISO 19111 [i.77] specifies a conceptual schema for the description of referencing by coordinates. It describes the minimum data required to define coordinate reference systems. The document supports the definition of: spatial coordinate reference systems where coordinate values do not change with time. It also describes how such a coordinate reference system can be related to a second coordinate reference system which could include time elements for referring to a temporal coordinate reference system.

Internet link to the standard: ISO 19111:2019

5.3.27 Geographic information: Referencing by geographic identifiers

ISO 19112 [i.80] specifies a general model for spatial referencing using geographic identifiers and defines the components of a spatial reference system. The specification bases on ISO 19111 [i.77] and adds a mechanism for recording complementary reference systems. The purpose of the document is to specify ways to define and describe systems of spatial references using geographic identifiers. It only covers the definition and recording of the referencing feature, and does not consider the forms of the relationship of the position relative to that feature.

Internet link to the standard: ISO 19112:2019

5.4 ITU

5.4.1 High Efficiency Video Coding (HEVC)

This video codec (Recommendation ITU-T H.265 [i.28]) splits the picture to be compressed into several block shaped regions and uses for each of these blocks inter- and intra-picture motion prediction and a two-dimensional transform technique for compression. The size of the prediction area is variable, different sizes are available and the best-fitting size with respect to compression ratio and resulting picture quality is chosen. The standard defines several profiles and allows the compression of videos for resolutions up to 8 k. The maximum frame rate to be supported is 300 frames per second.

Internet link to the standard: Recommendation ITU-T H.265

5.4.2 Versatile Video Coding (VVC)

This video codec (Recommendation ITU-T H.266 [i.29]) does not recommend any coding process for picture compression but defines a bitstream syntax that is variable to use. Any encoding process that is conform to the specified syntax fulfils the requirements of this standard. In this standard only the syntax format, semantics, and associated decoding process requirements are specified. Profiles, tiers and levels specify restrictions on bitstreams and limit the capabilities needed to decode the bitstreams allowing the implementation of decoders for different use cases.

Internet link to the standard: Recommendation ITU-T H.266

5.4.3 Digital Image Compression JPEG

This algorithm is optimized for the compression of digital images. The image is split into square-shaped regions and each of this region is transformed into a two-dimensional frequency domain. The coefficients of the values in the frequency domain are quantized according to the characteristics of the human psychovisual system and then transmitted. The compression methods used is usually lossy, as an option a lossless mode is also defined. Recommendation ITU-T T.81 [i.30] defines the coding algorithm and also specifies the file format for transmission.

Internet link to the standard: Recommendation ITU-T T.81

5.5 AOMedia

5.5.1 AOMedia Video 1 (AV1)

This video codec [i.31] splits the picture to be compressed into several block shaped regions and uses for each of these blocks inter- and intra-picture motion prediction and a two-dimensional transform technique for compression. There are several possibilities for partitioning the picture using square blocks, rectangle blocks or a combination of both. The best-fitting size with respect to compression ratio and resulting picture quality is chosen.

Internet link to the specification: AV1

5.6 United States Coast Guard

5.6.1 GNSS: Global Positioning System GPS

GPS [i.32] is a satellite-based navigation system to provide geolocation and timing information to receivers on or nearby the Earth. It is operated by the United States Government. For proper operation, the receiver needs line-of-sight connectivity to at least four satellites out of 24 that are in operation. By knowing the exact positions and the flight paths of the satellites and by processing the time differences of coded signals that are transmitted by these satellites, the receiver is able to calculate its own position. Under normal operating conditions the accuracy of the calculated position is in the range of a few meters. This accuracy can be improved into the range of centimetres by processing additionally to the satellite signals. signals that are transmitted by ground-based GPS systems the position of which is exactly known. This combination is named DGPS.

Internet link to a suite of reference documents: GPS

5.7 European Space Agency

5.7.1 GNSS: Galileo

Galileo [i.33] is a satellite-based navigation system to provide geolocation and timing information to receivers on or nearby the Earth. It is operated by the European Space Agency. For proper operation, the receiver needs line-of-sight connectivity to at least four satellites out of 24 that are in operation. By knowing the exact positions and the flight paths of the satellites and by processing the time differences of coded signals that are transmitted by these satellites, the receiver is able to calculate its own position. Under normal operating conditions the accuracy of the calculated position is in the range of one meter. This accuracy can be improved into the range of centimetres by combining several signal segments for the calculation of the receiver's position.

Internet link to a suite of articles and reference documents: Galileo

5.8 Bluetooth®

5.8.1 Bluetooth® Low Energy Indoor Positioning System

Devices that support Bluetooth® technologies can make use of the Low Energy radio technology of Bluetooth® to detect their position within a personal area network. To do so, locator beacons are deployed at known locations throughout the area of interest. The device enables the Bluetooth® radio and listens for locator beacons. Based on which locator beacons the device can receive, in combination with the received signal strength, the device is able to calculate its current position within the network. This technology provides accuracy down to the meter and centimetre level.

Internet link to the Bluetooth® Core Specification: for the physical layer description of the Low Energy profile, see Volume 6 of Bluetooth® Core Specification

5.9 IEEE

5.9.1 Wireless LAN Positioning

This technology makes use of processing and evaluating the propagation of radio waves between clients. For calculating the unknown position of a client, several technologies are existing. The technologies can be characterized by four classes:

1) Especially in an indoor environment, such a client can very often receive signals from several access points and can calculate the received signal strength for each of these points. Trilateration of the received signals gives the relative positioning of the receiving client that turns into absolute positioning if the positions of the access points is known.

2) A fingerprinting method uses the strength of the received signal from several access points and maps the received values to information provided by a database. A set of signals containing the received signal strengths is compared to values given in the database and the closest match is returned as the position of the client.

3) If the client device contains an antenna array or has access to it, the phase shift difference of the received signal at the antennas can be calculated and the angle of the received signal at each antenna leads to the calculation of the actual positioning of the client the position of which has to be estimated. Due to the processing power that is needed for this technology, the method can also be reversed. In such a case, the access points detect the angle of the incoming signal and transmit the result back to the client device.

4) By sending well-known data packets to a client and measuring the round trip delay of this transmission, the distance between two clients can be calculated. If this technology is used by incorporating several access points, a trilateration using all round trip delays leads to the actual position of the client.

All technologies mentioned in this clause can be combined to improve the estimation of the position (see IEEE 802.11az [i.35]).The accuracy for indoor positioning typically is in the range from a few decimetres to a few meters. The suite of IEEE 802.11 documents builds the base for the technologies described in this clause by specifying the physical layer and the possibilities for medium access control for wireless LAN.

Internet link to the IEEE 802 specification dealing with enhancements for positioning: IEEE 802.11az

5.9.2 Ultra Wide Band Positioning

This technology bases on the transmission of short data pulses that are spread over a wide frequency band. When using this technology for positioning, beacons are deployed at known locations and these beacons transmit identifiers and time stamps. By evaluating the time stamps, the run time of each pulse can be calculated and by assigning this run time to the appropriate beacon by the help of the identifier, the position of the receiving device can be calculated. This technology provides accuracy down to centimetre level (see IEEE 802.15.4z [i.36]).

Internet link to the physical layer description of the standard: IEEE 802.15.4z

5.10 Open Geospatial Consortium

5.10.1 OGC® SensorML: Model and XML Encoding Standard

This standard [i.37] provides a framework for defining processes and processing components associated with the measurement and post-measurement transformation of observations. It defines models and an XML schema to provide descriptions of sensors and to support the processing and the analysis of the sensor observations and the accompanying metadata. With the help of this framework the geometric, dynamic, and observational characteristics of sensors and sensor systems can be described.

Internet link to the standard: OGC® SensorML: Model and XML Encoding Standard

5.10.2 OGC® Sensor Things API Part 1: Sensing

This standard [i.38] defines a framework to interconnect Internet of Things devices, data, and applications over the web. A data model describes the elements of this framework and the requirements they have to fulfil. Access to sensors is given by APIs that allow IoT devices and applications to create, read, update and delete IoT data and metadata. Each access to a device is modelled as an observation that produces a result whose value is an estimate of a property of this device. The conformance with this standard can be checked by tests which are also part of the standard description.

Internet link to the standard: OGC® Sensor Things API Part 1: Sensing

5.10.3 OGC® Sensor Things API Part 2: Tasking Core

This standard [i.39] defines a framework to parameterize Internet of Things devices. A data model describes the elements of this framework and the requirements they have to fulfil. Access to sensors and their data is given by APIs that allow to create, read, update and delete IoT data and metadata. The conformance with this standard can be checked by tests which are also part of this document.

Internet link to the standard: OGC® Sensor Things API Part 2: Tasking Core

5.10.4 OGC® GeoPose

This standard [i.46] defines requirements for the interoperable exchange of the location and orientation of real or virtual geometric objects (poses) within reference frames that are anchored to the earth's surface or within other astronomical coordinate systems. An application can use GeoPose to position synthetic objects or their representations in the physical environment. The standard bases on an implementation-neutral model but it also gives examples how the elements of this standard can be represented in a JSON schema.

Internet link to the standard: OGC® GeoPose

5.10.5 OGC® Augmented Reality Markup Language

This standard [i.57] defines the Augmented Reality Markup Language 2.0 (ARML 2.0). ARML 2.0 allows users to describe virtual objects in an Augmented Reality (AR) scene with their appearances and their anchors (a broader concept of a location) related to the real world. Additionally, ARML 2.0 defines ECMAScript bindings to dynamically modify the AR scene based on user behaviour and user input. ARML can be serialized in XML or JSON and includes three concepts: features that represent a physical object that should be augmented, visual assets that describe the appearance of a virtual object in the augmented scene, and anchors that describe spatial relation between a physical and virtual object.

Internet link to the standard: OGC® ARML

5.10.6 OGC® Referencing by coordinates Corrigendum

This standard [i.78] contains editorial corrigenda to the specifications given in ISO 19111 [i.77] but also extends the set of definitions by specifying additional coordinate reference systems, defining the elements of these reference systems and giving requirements to describe coordinate metadata and the relationship between spatial coordinates.

Internet link to the standard: OGC® Referencing by Coordinates

5.10.7 OGC® API - Coordinate Reference Systems

This standard [i.79] defines the behaviour of a server that supports the ability to present geometry valued properties in a response document in one from a list of supported Coordinates Reference Systems (CRS). The standard assumes that each supported CRS can be referenced by a Uniform Resource Identifier (i.e. a URI). The document also specifies how a list of supported CRS is advertised and how the coordinates of geometry valued feature properties can be accessed in one of the supported CRSs.

Internet link to the standard: OGC® Coordinate Reference Systems

5.11 OASIS

5.11.1 Message Queuing Telemetry Transport (MQTT)

The standard [i.40] specifies a network protocol for machine-to-machine communication. It is especially designed for external applications with bandwidth restrictions and is event-driven. A message broker collects information from clients that are interested in offering their services (commonly called publishers) and from clients that are interested in receiving such services (commonly called subscribers). MQTT runs over any transport protocol that provides ordered, lossless, bi-directional connections. TCP/IP is the common connection protocol that is used. The exchange of information happens by pre-defined data packages called MQTT Control Packets that describe the type of data for this package, followed by the payload. The information is organized in a hierarchy of topics. The broker and the clients keep track of the session state of the information exchange between them.

Internet link to the standard: MQTT

5.12 American Society for Photogrammetry and Remote Sensing

5.12.1 LAS specification 1.4

The LiDAR technology calculates ranges by scanning objects or surfaces with laser pulses and measuring the time for the reflected pulses to return to the equipment. Most of the systems are airborne, but also terrestrial and mobile applications exist. Data is used to create high-resolution maps of the scanned environment, mainly represented as 3D point clouds. The technology itself is not standardized but there is a widely used file format for storing and transporting the scanned data. The LAS (Laser) specification issued by the ASPRS [i.41] describes the formats for point cloud data and for accompanying metadata.

Internet link to the standard: LAS 1.4

5.13 Khronos®

5.13.1 OpenXR™: Rendering

This specification [i.42] provides an API for developers to implement XR applications across a variety of different devices in a platform-agnostic way. Support is given for the instantiation of objects that store and track the state of an application and for controlling and monitoring the lifecycles of such objects. To present rendered images to the user in a controlled way, a storage model is used that allows communication between the application and the runtime environment of the device. Clause 10 of this specification [i.42] gives details about the supported characteristics of images as e.g. colour space, frame rate and format and how they are represented as elements of the interface.

Interlink link to the specification: OpenXR Specification, see clause 10

5.13.2 Collada™

Collada™ [i.43] defines an XML-based schema to facilitate the transport of 3D assets between applications - enabling diverse 3D authoring and content processing tools to be combined into a production pipeline. The intermediate language provides comprehensive encoding of visual scenes including geometry, shaders and effects, physics, animation, kinematics, and multiple version representations of the same asset. The specification defines an XML namespace and a database schema and thus supports all the features that 3D interactive authoring applications and digital content creation tools need to exchange and fully preserve asset data and metadata.

Internet link to the specification: Collada

5.13.3 Open XR™: Input Actions

This specification [i.42] provides an API for developers to implement XR applications across a variety of different devices in a platform-agnostic way. Support is given for the instantiation of objects that store and track the state of an application and for controlling and monitoring the lifecycles of such objects. The communication with input devices is done using so called actions. Actions are created at initialization time and later used to request input device state, create action spaces, or control haptic events. Input action handles represent actions that the application is interested in obtaining the state of. Direct input from a device hardware is not supported. Instead, an action is created that is then assigned to a specific input source. In such a way, the same action can be used for several input devices. Clause 11 of the specification [i.42] describes the structure of the interface for input actions and its elements.

Interlink link to the specification: OpenXR Specification, see clause 11

5.14 Wavefront Technologies

5.14.1 OBJ File Format

This file format [i.44] supports lines, polygons, and free-form curves and surfaces. Lines and polygons are described in terms of their points, while curves and surfaces are defined with control points and other information depending on the type of curve. The format supports rational and non-rational curves, including those based on Bezier, B-spline, Cardinal, and Taylor equations. The file format does not contain information about the texture and the colour of surfaces but it can reference such information that is stored in external libraries. Faces and surfaces can be assigned to named groups to facilitate their usage.

Internet link to the structure and the keywords of an OBJ file: OBJ File Format

5.15 Pixar® Animation Studios

5.15.1 Universal Scene Description (USD)

Universal Scene Description (USD) [i.60] is an open, extensible ecosystem originally created by Pixar Animation Studios for interchange of 3D computer graphics for describing, composing, simulating and collaboratively navigating and constructing 3D worlds. The format is binary or ascii encoded. Initially used in the field of VFX, the format enables advanced graphics features. Since more recently, the format is also used in industrial workflows with additional features dedicated to the IoT and digital twins.

USD does not only allow the authoring of 3D scenes but can also be used at runtime to interchange objects resp. to fill placeholders in the scene description by importing objects or by changing the attributes of already loaded objects.

Information about the format can be found here: USD

5.16 Autodesk

5.16.1 Autodesk Filmbox (FBX)

The Autodesk FBX [i.61] format is a closed, platform-independent 3D authoring and interchange format that provides access to 3D content from most 3D vendors. Available for Windows®, OSX, iOS and Linux operating systems, the free FBX Software Development Kit allows software and hardware vendors to easily add support for the FBX format. The SDK is a C++ software development toolkit that provides conversion utilities allowing application and content vendors to transfer their existing commercial and demonstration content into the FBX format. It includes advanced 3D graphics capabilities for creating meshes, animations, materials, etc.

Internet link to the details of the format: FBX

5.17 Apple®

5.17.1 ARKit

ARKit [i.66], introduced with iOS version 11, integrates hardware sensing features to produce augmented reality apps and games on Apple® devices (iPhone®, iPad®, Vision pro). ARKit combines device motion tracking, world tracking, scene understanding, face tracking, meshing and display conveniences to simplify building an AR experience.

Internet link to the details of the SDK: ARKit

5.17.2 RoomPlan

ARKit [i.66], introduced with iOS version 16, uses the camera and LiDAR Scanner on iPhone® and iPad® to create a 3D floor plan of a room, including key characteristics such as dimensions and types of furniture.

Internet link to the details of the API: RoomPlan

5.18 Google®

5.18.1 ARCore

ARCore [i.67] is Google®'s augmented reality SDK offering cross-platform APIs to build new immersive experiences on Android®, iOS, Unity, and Web. ARCore provides fundamental tools for building augmented reality experiences, including motion tracking, persistent anchors, face tracking, environmental understanding, depth sensing and light estimation. With the GeoSpatial API, it is possible to attach content remotely to any area mapped by Google® Street View and thus creating experiences on a global scale.

Internet link to the details of the SDK: ARCore

5.19 Qualcomm®

5.19.1 Snapdragon Spaces

Snapdragon Spaces [i.68] provides a uniform set of AR features designed to be independent of device manufacturers. Snapdragon Spaces is compatible with the Khronos® OpenXR™ specification [i.42] to enable application portability and is optimized for AR Glasses tethered to smartphones with a Khronos® OpenXR™ standards conformant runtime. The platform enables developers to build 3D applications for AR glasses from scratch or add headword AR features to existing Android smartphone applications to drive a unified, multiscreen experience between the smartphone screen in 2D and the real world in 3D. It includes advanced AR features such as motion tracking, scene understanding, persistent anchors , image and object recognition and tracking, hand tracking.

Internet link to the details of the SDK: Snapdragon Spaces

5.20 Cesium®

5.20.1 Cesium platform

Cesium is an open platform for creating 3D geospatial applications [i.81]. It enables the creation and retrieval of interactive 3D maps for visualizing complex geospatial data. It relies on the 3D tiles open specifications and can then be supplied with several types of 3D data. It comes with different implementations to be compatible with the main 3D ecosystems including Unity, Unreal and Omniverse.

Internet link to the details of the platform: Cesium Platform

5.20.2 3D Tiles

3D Tiles is an open specification [i.82] for sharing, visualizing, fusing, and interacting with massive heterogenous 3D geospatial content across desktop, web, and mobile applications. With a defined set of file formats, multiple types of 3D geospatial content including photogrammetry/massive models, BIM/CAD, 3D buildings, instanced features, and point clouds can be converted into 3D Tiles and combined into a single dataset.

Internet link to the details of the specifications: Cesium 3D Tiles

History

Document history		
V1.1.1	March 2025	Publication