International Standard

ISO/IEC 14496-1

# Information technology — Coding of audio-visual objects —

## Part 1: Systems

*Technologies de l'information — Codage des objets audiovisuels —*

*Partie 1: Systèmes*

Fifth edition
2026-07

Reference number
ISO/IEC 14496-1:2026(en)

© ISO/IEC 2026



**COPYRIGHT PROTECTED DOCUMENT**

© ISO/IEC 2026

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

# Contents

Page

**Foreword** ... **iv**
**Introduction** ... **v**
**1 Scope** ... **1**
**2 Normative references** ... **1**
**3 Terms and definitions** ... **2**
**4 Abbreviated terms** ... **9**
**5 Conventions** ... **10**
**6 Streaming framework** ... **10**
6.1 Systems decoder model ... 10
6.1.1 General ... 10
6.1.2 Concepts of the systems decoder model ... 11
6.1.3 Timing model specification ... 13
6.1.4 Buffer model specification ... 15
6.2 Object description framework ... 16
6.2.1 General ... 16
6.2.2 Common data structures ... 18
6.2.3 Intellectual property management and protection framework (IPMP) ... 21
6.2.4 Object content information (OCI) ... 23
6.2.5 Object descriptor stream ... 25
6.2.6 Object descriptor components ... 28
6.2.7 Rules for usage of the object description framework ... 60
6.2.8 Usage of the IPMP system interface ... 69
6.3 Synchronization of elementary streams ... 72
6.3.1 General ... 72
6.3.2 Sync layer ... 72
6.3.3 DMIF application interface ... 83
6.4 Multiplexing of elementary streams ... 83
6.4.1 General ... 83
6.4.2 M4Mux tool ... 83
6.4.3 M4Mux descriptors ... 89
**7 Profiles** ... **91**
**Annex A** (informative) **Time base reconstruction** ... **93**
**Annex B** (informative) **The QoS management model for ISO/IEC 14496 content** ... **96**
**Annex C** (informative) **Conversion between time and date conventions** ... **97**
**Annex D** (informative) **Graphical representation of object descriptor and sync layer syntax** ... **99**
**Annex E** (informative) **Elementary stream interface** ... **106**
**Annex F** (informative) **Upstream walkthrough** ... **108**
**Annex G** (informative) **Scene and object description carrousel** ... **113**
**Annex H** (normative) **Usage of ITU-T Recommendation H.264 | ISO/IEC 14496-10 AVC** ... **115**
**Bibliography** ... **118**

© ISO/IEC 2026 – All rights reserved

# Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This fifth edition cancels and replaces the fourth edition (ISO/IEC 14496-1:2010), which has been technically revised. It also incorporates the Amendments ISO/IEC 14496-1:2010/Amd.1:2010, ISO/IEC 14496-:2010/Amd.2:2014.

The main changes are as follows:

— Added support for LASeR

— Added support for raw audio and video bitstream

— Referencing to new Syntactic Description Language specification

A list of all parts in the ISO/IEC 14496 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

© ISO/IEC 2026 – All rights reserved

# Introduction

## 0.1 Overview

The ISO/IEC 14496 series specifies a system for the communication of interactive audio-visual scenes. This document includes the following elements:

a) the coded representation of natural or synthetic, two-dimensional (2D) or three-dimensional (3D) objects that can be manifested audibly and/or visually (audio-visual objects) (specified in ISO/IEC 14496-2, ISO/IEC 14496-3, ISO/IEC 14496-10, ISO/IEC 14496-11, ISO/IEC 14496-16, ISO/IEC 14496-19, ISO/IEC 14496-20, ISO/IEC 14496-23 and ISO/IEC 14496-25);

b) the coded representation of the spatio-temporal positioning of audio-visual objects as well as their behavior in response to interaction (scene description, specified in ISO/IEC 14496-11 ISO/IEC 14496-20);

c) the coded representation of information related to the management of data streams (synchronization, identification, description and association of stream content, specified in this document and in ISO/IEC 14496-24);

d) a generic interface to the data stream delivery layer functionality (specified in ISO/IEC 14496-6);

e) an application engine for programmatic control of the player: format, delivery of downloadable Java byte code as well as its execution lifecycle and behavior through APIs (specified in ISO/IEC 14496-11 and ISO/IEC 14496-21);

f) a file format to contain the media information of an ISO/IEC 14496 presentation in a flexible, extensible format to facilitate interchange, management, editing, and presentation of the media specified in ISO/IEC 14496-12 (ISO File Format), ISO/IEC 14496-14 (MP4 File Format) and ISO/IEC 14496-15 (AVC File Format); and

g) the coded representation of font data and of information related to the management of text streams and font data streams (specified in ISO/IEC 14496-17, ISO/IEC 14496-18 and ISO/IEC 14496-22).

The overall operation of a system communicating audio-visual scenes can be paraphrased as follows:

At the sending terminal, the audio-visual scene information is compressed, supplemented with synchronization information and passed to a delivery layer that multiplexes it into one or more coded binary streams that are transmitted or stored. At the receiving terminal, these streams are demultiplexed and decompressed. The audio-visual objects are composed according to the scene description and synchronization information and presented to the end user. The end user may have the option to interact with this presentation. Interaction information can be processed locally or transmitted back to the sending terminal. ISO/IEC 14496 defines the syntax and semantics of the bitstreams that convey such scene information, as well as the details of their decoding processes.

This document specifies the following tools:

— a terminal model for time and buffer management;

— a coded representation of metadata for the identification, description and logical dependencies of the elementary streams (object descriptors and other descriptors);

— a coded representation of descriptive audio-visual content information (object content information – OCI);

— an interface to intellectual property management and protection (IPMP) systems;

— a coded representation of synchronization information (sync layer – SL) and

— a multiplexed representation of individual elementary streams in a single stream (M4Mux).

These various elements are described functionally in this Subclause and specified in the normative Clauses that follow.

© ISO/IEC 2026 – All rights reserved

## 0.2 Architecture

The information representation specified in the ISO/IEC 14496 series describes the means to create an interactive audio-visual scene in terms of coded audio-visual information and associated scene description information. The entity that composes and sends or receives and presents such a coded representation of an interactive audio-visual scene is generically referred to as an "audio-visual terminal" or just "terminal". This terminal may correspond to a standalone application or be part of an application system.

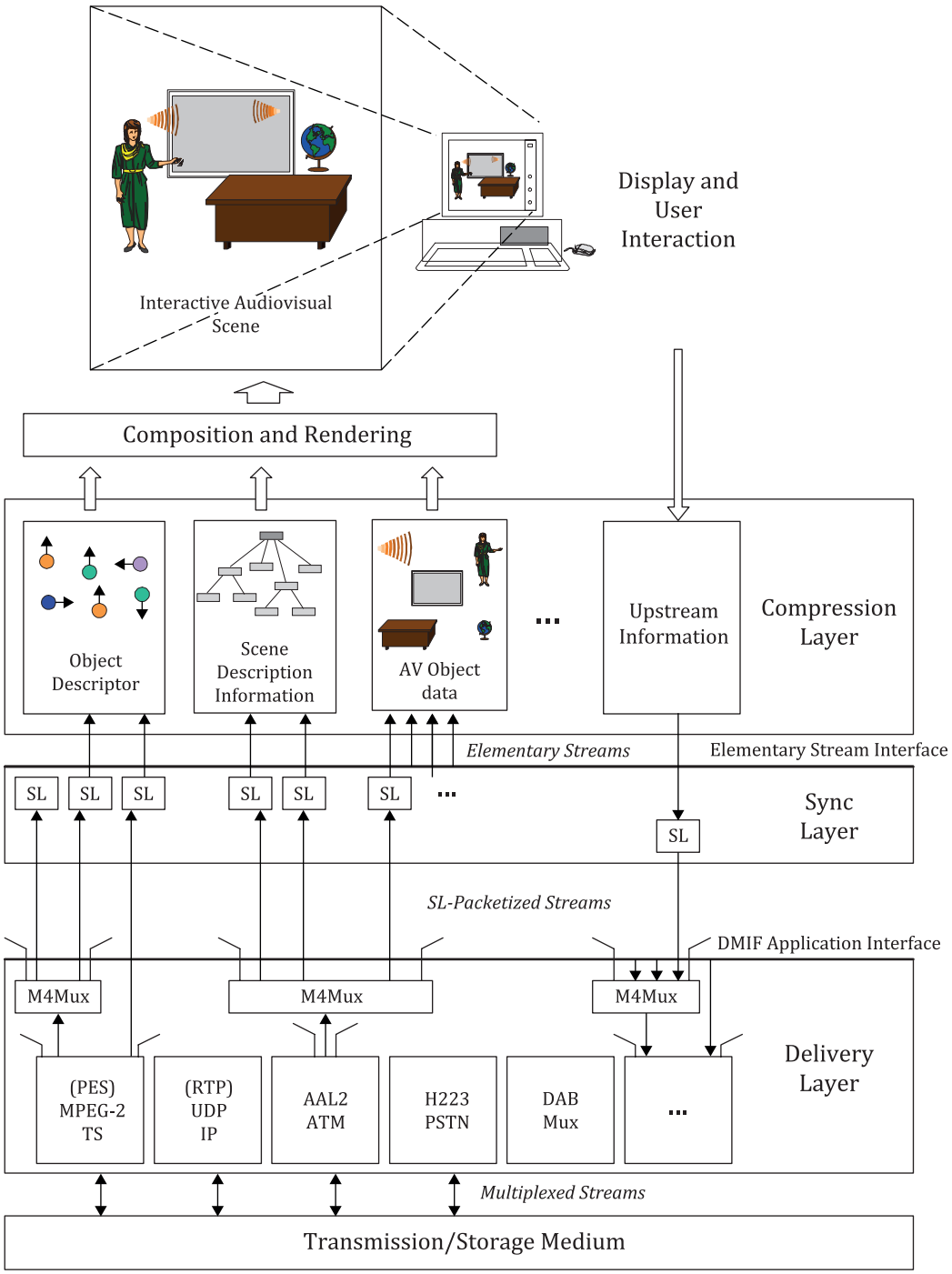



**Figure 1 — The ISO/IEC 14496 series terminal architecture**

The basic operations performed by such a receiver terminal are as follows. Information that allows access to content complying with the ISO/IEC 14496 series is provided as initial session set up information to the terminal. ISO/IEC 14496-6 defines the procedures for establishing such session contexts as well as the interface to the delivery layer that generically abstracts the storage or transport medium. The initial set

© ISO/IEC 2026 – All rights reserved

up information allows, in a recursive manner, to locate one or more elementary streams that are part of the coded content representation. Some of these elementary streams may be grouped together using the multiplexing tool described in this document.

Elementary streams contain the coded representation of either audio or visual data or scene description information or user interaction data or text or font data. Elementary streams may as well themselves convey information to identify streams, to describe logical dependencies between streams, or to describe information related to the content of the streams. Each elementary stream contains only one type of data.

Elementary streams are decoded using their respective stream-specific decoders. The audio-visual objects are composed according to the scene description information and presented by the terminal's presentation device(s). All these processes are synchronized according to the systems decoder model (SDM) using the synchronization information provided at the synchronization layer.

These basic operations are depicted in Figure 1 and are described in more detail below.

### 0.3 Terminal model: systems decoder model

The systems decoder model provides an abstract view of the behavior of a terminal complying with this document. Its purpose is to enable a sending terminal to predict how the receiving terminal will behave in terms of buffer management and synchronization when reconstructing the audio-visual information that comprises the presentation. The systems decoder model includes a systems timing model and a systems buffer model which are described briefly in the following Subclauses.

#### 0.3.1 Timing model

The timing model defines the mechanisms through which a receiving terminal establishes a notion of time that enables it to process time-dependent events. This model also allows the receiving terminal to establish mechanisms to maintain synchronization both across and within particular audio-visual objects as well as with user interaction events. In order to facilitate these functions at the receiving terminal, the timing model requires that the transmitted data streams contain implicit or explicit timing information. Two sets of timing information are defined in this document: clock references and time stamps. The former convey the sending terminal's time base to the receiving terminal, while the latter convey a notion of relative time for specific events such as the desired decoding or composition time for portions of the encoded audio-visual information.

#### 0.3.2 Buffer model

The buffer model enables the sending terminal to monitor and control the buffer resources that are needed to decode each elementary stream in a presentation. The required buffer resources are conveyed to the receiving terminal by means of descriptors at the beginning of the presentation. The terminal can then decide whether or not it is capable of handling this particular presentation. The buffer model allows the sending terminal to specify when information may be removed from these buffers and enables it to schedule data transmission so that the appropriate buffers at the receiving terminal do not overflow or underflow.

### 0.4 Multiplexing of streams: the delivery layer

The term delivery layer is used as a generic abstraction of any existing transport protocol stack that may be used to transmit and/or store content complying with ISO/IEC 14496. The functionality of this layer is not within the scope of this document, and only the interface to this layer is considered. This interface is the DMIF Application Interface (DAI) specified in ISO/IEC 14496-6. The DAI defines not only an interface for the delivery of streaming data, but also for signaling information required for session and channel set up as well as tear down. A wide variety of delivery mechanisms exist below this interface, with some of them indicated in Figure 1. These mechanisms serve for transmission as well as storage of streaming data, i.e. a file is considered to be a particular instance of a delivery layer. For applications where the desired transport facility does not fully address the needs of a service according to the specifications in ISO/IEC 14496, a simple multiplexing tool (M4Mux) with low delay and low overhead is defined in this document.

### 0.5 Synchronization of streams: the sync layer

Elementary streams are the basic abstraction for any streaming data source. Elementary streams are conveyed as sync layer-packetized (SL-packetized) streams at the DMIF Application Interface. This

© ISO/IEC 2026 – All rights reserved

packetized representation additionally provides timing and synchronization information, as well as fragmentation and random access information. The sync layer (SL) extracts this timing information to enable synchronized decoding and, subsequently, composition of the elementary stream data.

### 0.6 The compression Layer

The compression layer receives data in its encoded format and performs the necessary operations to decode this data. The decoded information is then used by the terminal's composition, rendering and presentation subsystems.

#### 0.6.1 Object description framework

The purpose of the object description framework is to identify and describe elementary streams and to associate them appropriately to an audio-visual scene description. Object descriptors serve to gain access to ISO/IEC 14496 content. Object content information and the interface to intellectual property management and protection systems are also part of this framework.

An object descriptor is a collection of one or more elementary stream descriptors that provide the configuration and other information for the streams that relate to either an audio-visual object, or text or font data, or a scene description. Object descriptors are themselves conveyed in elementary streams. Each object descriptor is assigned an identifier (object descriptor ID), which is unique within a defined name scope. This identifier is used to associate audio-visual objects in the scene description with a particular object descriptor, and thus the elementary streams related to that particular object.

Elementary stream descriptors include information about the source of the stream data, in form of a unique numeric identifier (the elementary stream ID) or a URL pointing to a remote source for the stream. Elementary stream descriptors also include information about the encoding format, configuration information for the decoding process and the sync layer packetization, as well as quality of service requirements for the transmission of the stream and intellectual property identification. Dependencies between streams can also be signaled within the elementary stream descriptors. This functionality may be used, for example, in scalable audio or visual object representations to indicate the logical dependency of a stream containing enhancement information, to a stream containing the base information. It can also be used to describe alternative representations for the same content (e.g. the same speech content in various languages).

##### 0.6.1.1 Intellectual property management and protection

The intellectual property management and protection (IPMP) framework for ISO/IEC 14496 content consists of a normative interface that permits an ISO/IEC 14496 terminal to host one or more IPMP systems in the form of monolithic IPMP systems or modular IPMP tools. The IPMP interface consists of IPMP elementary streams and IPMP descriptors. IPMP descriptors are carried as part of an object descriptor stream. IPMP elementary streams carry time variant IPMP information that can be associated to multiple object descriptors.

The IPMP system, or IPMP tools themselves are non-normative components that provides intellectual property management and protection functions for the terminal. The IPMP systems or tools uses the information carried by the IPMP elementary streams and descriptors to make protected ISO/IEC 14496 content available to the terminal.

The intellectual property management and protection (IPMP) framework for ISO/IEC 14496 content consists of a set of tools that permits an ISO/IEC 14496 terminal to support IPMP functionality. This functionality is provided by two different complementary technologies, supporting different levels of interoperability:

— The IPMP framework as defined in 6.2.3, consists of a normative interface that permits an ISO/IEC 14496 terminal to host one or more IPMP systems. The IPMP interface consists of IPMP elementary streams and IPMP descriptors. IPMP descriptors are carried as part of an object descriptor stream. IPMP elementary streams carry time variant IPMP information that can be associated to multiple object descriptors. The IPMP system itself is a non-normative component that provides intellectual property management and protection functions for the terminal. The IPMP system uses the information carried by the IPMP elementary streams and descriptors to make protected ISO/IEC 14496 content available to the terminal.

© ISO/IEC 2026 – All rights reserved

— The IPMP framework extension, as specified in ISO/IEC 14496-13 allows, in addition to the functionality specified in ISO/IEC 14496-1, a finer granularity of governance. ISO/IEC 14496-13 provides normative support for individual IPMP components, referred to as IPMP tools, to be normatively placed at identified points of control within the terminal systems model. Additionally, ISO/IEC 14496-13 provides normative support for secure communications to be performed between IPMP tools. ISO/IEC 14496-1 also specifies specific normative extensions at the systems level to support the IPMP functionality described in ISO/IEC 14496-13.

An application may choose not to use an IPMP system, thereby offering no management and protection features.

#### 0.6.1.2 Object content information

Object content information (OCI) descriptors convey descriptive information about audio-visual objects. The main content descriptors are: content classification descriptors, keyword descriptors, rating descriptors, language descriptors, textual descriptors, and descriptors about the creation of the content. OCI descriptors can be included directly in the related object descriptor or elementary stream descriptor or, if it is time variant, it may be carried in an elementary stream by itself. An OCI stream is organized in a sequence of small, synchronized entities called events that contain a set of OCI descriptors. OCI streams can be associated to multiple object descriptors.

### 0.6.2 Scene description streams

Scene description addresses the organization of audio-visual objects in a scene, in terms of both spatial and temporal attributes. This information allows the composition and rendering of individual audio-visual objects after the respective decoders have reconstructed the streaming data for them. For visual data, ISO/IEC 14496-11 does not mandate particular composition algorithms. Hence, visual composition is implementation dependent. For audio data, the composition process is defined in a normative manner in ISO/IEC 14496-11 and ISO/IEC 14496-3.

The scene description is represented using a parametric approach (BIFS - dinary format for scenes). The description consists of an encoded hierarchy (tree) of nodes with attributes and other information (including event sources and targets). Leaf nodes in this tree correspond to elementary audio-visual data, whereas intermediate nodes group this material to form audio-visual objects, and perform grouping, transformation, and other such operations on audio-visual objects (scene description nodes). The scene description can evolve over time by using scene description updates.

In order to facilitate active user involvement with the presented audio-visual information, ISO/IEC 14496-11 provides support for user and object interactions. Interactivity mechanisms are integrated with the scene description information, in the form of linked event sources and targets (routes) as well as sensors (special nodes that can trigger events based on specific conditions). These event sources and targets are part of scene description nodes, and thus allow close coupling of dynamic and interactive behavior with the specific scene at hand. ISO/IEC 14496-11, however, does not specify a particular user interface or a mechanism that maps user actions (e.g. keyboard key presses or mouse movements) to such events.

Such an interactive environment may not need an upstream channel, but ISO/IEC 14496 also provides means for client-server interactive sessions with the ability to set up upstream elementary streams and associate them to specific downstream elementary streams.

### 0.6.3 Audio-visual streams

The coded representation of audio and visual information are described in ISO/IEC 14496-3 (Audio) and ISO/IEC 14496-2 (Visual) and ISO/IEC 14496-10 (Advanced Video Coding) respectively. The reconstructed audio-visual data are made available to the composition process for potential use during the scene rendering.

### 0.6.4 Upchannel streams

Downchannel elementary streams may require upchannel information to be transmitted from the receiving terminal to the sending terminal (e.g. to allow for client-server interactivity). Figure 1 indicates the flowpath for an elementary stream from the receiving terminal to the sending terminal. The content of upchannel streams is specified in the same part of the specification that defines the content of the downstream

© ISO/IEC 2026 – All rights reserved

data. For example, upchannel control streams for video downchannel elementary streams are defined in ISO/IEC 14496-2.

### 0.6.5 Interaction streams

The coded representation of user interaction information is not in the scope of ISO/IEC 14496. But this information is to be translated into scene modification and the modifications made available to the composition process for potential use during the scene rendering.

### 0.6.6 Text and font data streams

Scene description often contains information presented in textual format. The audio-visual data encoded in the scene may also be accompanied by supplemental text information such as subtitles. In order to enable time-based updates of text data and to insure the text appearance and layout, both elementary streams carrying timed text information and font data are used. The coded representation of the timed text stream is described in ISO/IEC 14496-17. The font data format and encoded representation of font data stream are described in ISO/IEC 14496-18 (font data stream) and ISO/IEC 14496-22 (font data format).

## 0.7 Application engine

The MPEG-J is a programmatic system (as opposed to a conventional parametric system) which specifies API(s) for interoperation of MPEG-4 media players with Java code. By combining MPEG-4 media and safe executable code, content creators may embed complex control and data processing mechanisms with their media data to intelligently manage the operation of the audio-visual session. The parametric MPEG-4 system forms the presentation engine while the MPEG-J subsystem controlling the presentation engine forms the application engine.

The Java application is delivered as a separate elementary stream to the MPEG-4 terminal. There it will be directed to the MPEG-J run time environment, from where the MPEG-J program will have access to the various components and required data of the MPEG-4 player to control it.

In addition to the basic packages of the language (java.lang, java.io, java.util) a few categories of APIs have been defined for different scopes. For the scene graph API the objective is to provide access to the scene graph specified in ISO/IEC 14496-11: to inspect the graph, to alter nodes and their fields, and to add and remove nodes within the graph. The resource API is used for regulation of performance: it provides a centralized facility for managing resources. This is used when the program execution is contingent upon the terminal configuration and its capabilities, both static (that do not change during execution) and dynamic. Decoder API allows the control of the decoders that are present in the terminal. The net API provides a way to interact with the network, being compliant to the MPEG-4 DMIF application interface. Complex applications and enhanced interactivity are possible with these basic packages. The architecture of MPEG-J is presented in more detail in ISO/IEC 14496-11.

## 0.8 Extensible MPEG-4 textual (XMT) format

The extensible MPEG-4 textual (XMT) format is a textual representation of the multimedia content described in ISO/IEC 14496 using the extensible markup language (XML). XMT is designed to facilitate the creation and maintenance of MPEG-4 multimedia content, whether by human authors or by automated machine programs. XMT is specified in ISO/IEC 14496-11.

The textual representation of MPEG-4 content has high-level abstractions, XMT-O, that allow authors to exchange their content easily with other authors or authoring tools, while at the same time preserving semantic intent. XMT also has low-level textual representations, XMT-A, covering the full scope and function of MPEG-4. The high-level XMT-O is designed to facilitate interoperability with the synchronized multimedia integration Language (SMIL) 2.0, a recommendation from the W3C consortium, and also with extensible 3D specification, X3D, developed by the Web3D consortium as the next generation of virtual reality modeling language (VRML).

The XMT language has grammars that are specified using the W3C XML schema language. The grammars contain rules for element placement and attribute values, etc. These rules for XMT, defined using the schema language, follow the binary coding rules defined in ISO/IEC 14496-11 and help ensure that the textual representation can be coded into correct binary according to ISO/IEC 14496-11 coding rules.

© ISO/IEC 2026 – All rights reserved

All constructs in the ISO/IEC 14496 specification have their parallel in the XMT textual format. For the visual and audio parts, XMT provides a means to reference external media streams of either pre-encoded or raw audiovisual binary content. While XMT does not contain a textual format for audiovisual media, it does contain hints in a textual format that allow an XMT tool to encode and embed the audiovisual media into a complete MPEG-4 presentation.

© ISO/IEC 2026 – All rights reserved

# Information technology — Coding of audio-visual objects —

# Part 1: Systems

## 1 Scope

This document specifies system level functionalities for the communication of interactive audio-visual scenes, i.e. the coded representation of information related to the management of data streams (synchronization, identification, description and association of stream content).

## 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 639:2023, *Code for individual languages and language groups*

ISO/IEC 10646:2020, *Information technology — Universal coded character set (UCS)*

ISO/IEC 10918-1:1994 | ITU-T Rec. T.81, *Information technology — Digital compression and coding of continuous-tone still images — Part 1: Requirements and guidelines*

ISO/IEC 11172-2:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 2: Video*

ISO/IEC 11172-3:1993, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio*

ISO/IEC 13818-2:2000 | ITU-T Rec. H.262, *Information technology — Generic coding of moving pictures and associated audio information — Part 2: Video*

ISO/IEC 13818-3:1998, *Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio*

ISO/IEC 13818-7:2006, *Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC)*

ISO/IEC 14496-2:2004, *Information technology — Coding of audio-visual objects — Part 2: Visual*

ISO/IEC 14496-3:2019, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-10:2025, *Information technology — Coding of audio-visual objects — Part 10: Advanced video coding*

ISO/IEC 14496-15:2024, *Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format*

ISO/IEC 14496-16:2011, *Information technology — Coding of audio-visual objects — Part 16: Animation Framework eXtension (AFX)*

ISO/IEC 14496-18:2004, *Information technology — Coding of audio-visual objects — Part 18: Font compression and streaming*

© ISO/IEC 2026 – All rights reserved

ISO/IEC 14496-34:2025, *Information technology — Coding of audio-visual objects — Part 34: Syntactic description language*

DAVIC 1.4.1 specification, *Part 9: Information Representation*

ANSI/SMPTE 291M-1996, *Television — Ancillary Data Packet and Space Formatting*

SMPTE 315M-1999, *Television — Camera Positioning Information Conveyed by Ancillary Data Packets*

W3C Recommendation: August 2001 — *Synchronized Multimedia Integration Language (SMIL 2.0)*, https://www.w3.org/TR/smil20/

W3C Recommendation: May 2001 — *XML Schema*, https://www.w3.org/TR/xmlschema-0/

## 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

**3.1**
**access unit**
**AU**
smallest individually accessible portion of data within an *elementary stream* (3.27) to which unique timing information can be attributed

**3.2**
**audio-visual object**
representation of a natural or synthetic object that has an audio and/or visual manifestation

Note 1 to entry: The representation corresponds to a node or a group of nodes in the BIFS scene description. Each audio-visual object is associated with zero or more *elementary streams* using one or more *object descriptors.*

**3.3**
**audio-visual scene**
**AV Scene**
set of audio-visual objects together with scene description information that defines their spatial and temporal attributes including behaviors resulting from object and user interactions

**3.4**
**AVC parameter set**
sequence parameter set or a picture parameter set

**3.5**
**AVC access unit**
access unit made up of NAL Units as defined in ITU-T H.264 | ISO/IEC 14496-10 with the structure defined in 5.2.3 of ISO/IEC 14496-15:2024

**3.6**
**AVC parameter set access unit**
access unit made up only of sequence parameter set NAL units or picture parameter set NAL units having same timestamps to be applied

**3.7**
**AVC parameter set elementary stream**
elementary stream containing made up only of AVC parameter set access units

© ISO/IEC 2026 – All rights reserved

**3.8**
**AVC video elementary stream**
elementary stream containing access units made up of NAL units for coded picture data

**3.9**
**binary format for scene**
**BIFS**
coded representation of a parametric scene description format as specified in ISO/IEC 14496-11

**3.10**
**buffer model**
model that defines how a terminal complying with ISO/IEC 14496 manages the buffer resources that are needed to decode a presentation

**3.11**
**clock reference**
special time stamp that conveys a reading of a time base

**3.12**
**composition**
process of applying scene description information in order to identify the spatio-temporal attributes and hierarchies of audio-visual objects

**3.13**
**composition memory**
**CM**
random access memory that contains composition units

**3.14**
**composition time stamp**
**CTS**
indication of the nominal composition time of a composition unit

**3.15**
**composition unit**
**CU**
individually accessible portion of the output that a decoder produces from access units

**3.16**
**compression layer**
layer of a system according to the specifications in ISO/IEC 14496 that translates between the coded representation of an elementary stream and its decoded representation. It incorporates the decoders

**3.17**
**control point**
point on a given elementary stream in a terminal where IPMP processing on stream data shall be carried out

**3.18**
**decoder**
entity that translates between the coded representation of an elementary stream and its decoded representation

**3.19**
**decoding buffer**
**DB**
buffer at the input of a decoder that contains access units

**3.20**
**decoder configuration**
configuration of a decoder for processing its elementary stream data by using information contained in its elementary stream descriptor

© ISO/IEC 2026 – All rights reserved

**3.21**
**decoding time stamp**
**DTS**
indication of the nominal decoding time of an access unit

**3.22**
**delivery layer**
generic abstraction for delivery mechanisms (computer networks, etc.) able to store or transmit a number of multiplexed elementary streams or M4Mux streams

**3.23**
**descriptor**
data structure that is used to describe particular aspects of an elementary stream or a coded audio-visual object

**3.24**
**DMIF application interface**
**DAI**
interface specified in ISO/IEC 14496-6 used to model the exchange of SL-packetized stream data and associated control information between the sync layer and the delivery layer

**3.25**
**elementary stream**
**ES**
consecutive flow of mono-media data from a single source entity to a single destination entity on the compression layer

**3.26**
**elementary stream descriptor**
structure contained in object descriptors that describes the encoding format, initialization information, sync layer configuration, and other descriptive information about the content carried in an elementary stream

**3.27**
**elementary stream interface**
**ESI**
conceptual interface modeling the exchange of elementary stream data and associated control information between the compression layer and the sync layer

**3.28**
**M4Mux channel**
**FMC**
label to differentiate between data belonging to different constituent streams within one M4Mux stream

Note 1 to entry: A sequence of data in one M4Mux channel within a M4Mux stream corresponds to one single SL-packetized stream.

**3.29**
**M4Mux packet**
smallest data entity managed by the M4Mux tool consisting of a header and a payload

**3.30**
**M4Mux stream**
sequence of M4Mux packets with data from one or more SL-packetized streams that are each identified by their own M4Mux channel

**3.31**
**M4Mux tool**
tool that allows the interleaving of data from multiple data streams

© ISO/IEC 2026 – All rights reserved

**3.32**
**graphics profile**
profile that specifies the permissible set of graphical elements of the BIFS tool that may be used in a scene description stream

Note 1 to entry: BIFS comprises both graphical and scene description elements.

**3.33**
**inter**
mode for coding parameters that uses previously coded parameters to construct a prediction

**3.34**
**interaction stream**
elementary stream that conveys user interaction information

**3.35**
**intra**
mode for coding parameters that does not make reference to previously coded parameters to perform the encoding

**3.36**
**initial object descriptor**
special object descriptor that allows the receiving terminal to gain initial access to portions of content encoded according to ISO/IEC 14496 and that conveys profile and level information to describe the complexity of the content

**3.37**
**intellectual property identification**
**IPI**
unique identification of one or more elementary streams corresponding to parts of one or more audio-visual objects

**3.38**
**intellectual property management and protection system**
generic term for mechanisms and tools to manage and protect intellectual property

Note 1 to entry: This document defines the interface to such systems as well as:

— The provision for the identification of IPMP tools through the use of a functional description of the IPMP tools' capabilities in a parametric fashion.

— Controlling the time of instantiation of IPMP tools either by the inclusion of references to the required IPMP tools or at the request of already instantiated IPMP tools.

— Providing secure messaging between IPMP tools and the terminal and between IPMP tools and the user.

— Notification of the instantiation of IPMP tools to IPMP tools requesting such notification.

— Interaction between IPMP tools, and/or the terminal and the user.

— The carriage of IPMP tools within the bitstream.

**3.39**
**IPMP information**
Information directed to a given IPMP tool to enable, assist or facilitate its operation

© ISO/IEC 2026 – All rights reserved

**3.40**
**IPMP system**
monolithic IPMP protection scheme which requires implementation dependant access to protected streams at required control points and must provide any intra-communication within the system on an implementation basis

Note 1 to entry: In this document, the term "IPMP system" is used in some cases to indicate either an actual IPMP system or a combination of IPMP tools whose combination provides the functionality of an IPMP system. In cases where the distinction is important the proper respective terms are used.

**3.41**
**IPMP tool**
modules that perform (one or more) IPMP functions such as authentication, decryption, watermarking, etc.

Note 1 to entry: Conceptually the use of one or more IPMP tools is combined to perform the functionality of an IPMP system. IPMP tools, as opposed to IPMP systems, are normatively identified as to which control points they function at as well as are provided normative methods for secure communications both within as well as outside of a given IPMP tools comprised functional "IPMP system". An additional difference between IPMP tools and IPMP systems is that IPMP tools, or a combination thereof, may be used for the protection of object streams.

**3.42**
**IPMP tool identifier**
unambiguous identifier for IPMP tools at the presentation level or at a universal level

Note 1 to entry: Two different identifiers are provided to support the differentiation between the use of IPMP systems and IPMP tools.

**3.43**
**IPMP tool list**
list of selectable IPMP tool list required to process the Content

**3.44**
**media node**
time dependent BIFS nodes that refer to a media stream through a URL field in the following list: AnimationStream, AudioBuffer, AudioClip, AudioSource, Inline, MovieTexture

**3.45**
**media stream**
one or more elementary streams whose ES descriptors are aggregated in one object descriptor and that are jointly decoded to form a representation of an AV object

**3.46**
**media time line**
time line expressing normal play back time of a media stream

**3.47**
**MP4 file**
name of the file format described in ISO/IEC 14496-14

**3.48**
**object clock reference**
**OCR**
clock reference that is used by a decoder to recover the time base of the encoder of an elementary stream

**3.49**
**object content information**
**OCI**
additional information about content conveyed through one or more elementary streams; either aggregated to individual elementary stream descriptors or is itself conveyed as an elementary stream.

© ISO/IEC 2026 – All rights reserved

**3.50**
**object descriptor**
**OD**
descriptor that aggregates one or more elementary streams by means of their elementary stream descriptors and defines their logical dependencies

**3.51**
**object descriptor command**
command that identifies the action to be taken on a list of object descriptors or object descriptor IDs, e.g. update or remove

**3.52**
**object descriptor profile**
profile that specifies the configurations of the object descriptor tool and the sync layer tool that are allowed

**3.53**
**object descriptor stream**
elementary stream that conveys object descriptors encapsulated in object descriptor commands

**3.54**
**object time base**
**OTB**
time base valid for a given elementary stream, and hence for its decoder; conveyed to the decoder via object clock references and which is used by all time stamps relating to this object's decoding process

**3.55**
**parametric audio decoder**
set of tools for representing and decoding speech signals coded at bit rates between 6 Kbit/s and 16 Kbit/s, according to the specifications in ISO/IEC 14496-3

**3.56**
**parametric description**
SDL declaration that describes the parametric configuration and other interface message(s) that drive the tool and the behaviour defined for fulfilment of such a description

**3.57**
**quality of service**
**QoS**
performance that an elementary stream requests from the delivery channel through which it is transported

Note 1 to entry: QoS is characterized by a set of parameters (e.g. bit rate, delay jitter, bit error rate, etc.).

**3.58**
**random access**
process of beginning to read and decode a coded representation at an arbitrary point within the elementary stream

**3.59**
**reference point**
location in the data or control flow of a system that has some defined characteristics

**3.60**
**rendering**
action of transforming a scene description and its constituent audio-visual objects from a common representation space to a specific presentation device (i.e. speakers and a viewing window)

**3.61**
**rendering area**
portion of the display device's screen into which the scene description and its constituent audio-visual objects are to be rendered

© ISO/IEC 2026 – All rights reserved

**3.62**
**scene description**
information that describes the spatio-temporal positioning of audio-visual objects as well as their behavior resulting from object and user interactions and which makes reference to elementary streams with audio-visual data by means of pointers to object descriptors

**3.63**
**scene description stream**
elementary stream that conveys scene description information

**3.64**
**scene graph element**
element of the BIFS language that relates only to the structure of the audio-visual scene (spatio-temporal positioning of audio-visual objects as well as their behavior resulting from object and user interactions) excluding the audio, visual and graphics nodes as specified in ISO/IEC 14496-11

**3.65**
**scene graph profile**
profile that defines the permissible set of scene graph elements of the BIFS tool that may be used in a scene description stream

Note 1 to entry: BIFS comprises both graphical and scene description elements.

**3.66**
**seekable**
property of a media stream for which the play back is possible from any position

**3.67**
**SL-packetized stream**
**SPS**
sequence of sync layer packets that encapsulate one elementary stream

**3.68**
**stream object**
media stream or a segment thereof, referenced through a URL field in the scene in the form “OD:n” or “OD:n#<segmentName>”

**3.69**
**structured audio**
method of describing synthetic sound effects and music as defined by ISO/IEC 14496-3

**3.70**
**sync layer**
**SL**
layer to adapt elementary stream data for communication across the DMIF application interface, providing timing and synchronization information, as well as fragmentation and random access information

Note 1 to entry: The sync layer syntax is configurable and can be configured to be empty.

**3.71**
**sync layer configuration**
configuration of the sync layer syntax for a particular elementary stream using information contained in its elementary stream descriptor

**3.72**
**sync layer packet**
**SL-Packet**
smallest data entity managed by the sync layer consisting of a configurable header and a payload which may consist of one complete access unit or a partial access unit

© ISO/IEC 2026 – All rights reserved

**3.73**
**systems decoder model**
**SDM**
model that provides an abstract view of the behavior of a terminal compliant to ISO/IEC 14496 which consists of the buffer model and the timing model

**3.74**
**system time base**
**STB**
time base of the terminal whose resolution is implementation-dependent and according to which all operations in the terminal are performed

**3.75**
**terminal**
system that sends, or receives and presents the coded representation of an interactive audio-visual scene as defined by ISO/IEC 14496-11 which can be a standalone system, or part of an application system complying with ISO/IEC 14496

**3.76**
**time base**
clock, equivalent to a counter that is periodically incremented

**3.77**
**timing model**
model that specifies the semantic meaning of timing information, how it is incorporated (explicitly or implicitly) in the coded representation of information, and how it can be recovered at the receiving terminal

**3.78**
**time stamp**
indication of a particular time instant relative to a time base

**3.79**
**track**
collection of related samples in an MP4 file

## 4 Abbreviated terms

| | |
|---|---|
| AU | access unit |
| AV | audio-visual |
| AVC | advanced video coding, ITU-T Recommendation H.264 \| ISO/IEC 14496-10 |
| BIFS | binary format for scene |
| CM | composition memory |
| CTS | composition time stamp |
| CU | composition unit |
| DAI | DMIF application interface (see ISO/IEC 14496-6) |
| DB | decoding buffer |
| DTS | decoding time stamp |
| ES | elementary stream |
| ESI | elementary stream Interface |

© ISO/IEC 2026 – All rights reserved