



**International
Standard**

ISO/IEC 14496-12

**Information technology — Coding of
audio-visual objects —**

**Part 12:
ISO base media file format**

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

Partie 12: Format ISO de base pour les fichiers médias

**Eighth edition
2026-04**

Sample Document

get full document from standards.iteh.ai



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

© ISO/IEC 2026 – All rights reserved

Contents

	Page
Foreword	x
Introduction	xi
1 Scope	1
2 Normative references	1
3 Terms, definitions and abbreviated terms	2
3.1 Terms and definitions.....	2
3.2 Abbreviated terms.....	8
4 ISO base media file organization	9
4.1 File structure.....	9
4.2 Binary structure.....	9
4.2.1 Syntax conventions.....	9
4.2.2 Box definitions.....	10
4.2.3 Extensibility of box definitions.....	11
4.2.4 Data types and fields.....	12
5 Typed ISO base media file organization	13
5.1 File structure.....	13
5.2 File-type box.....	14
5.2.1 Definition.....	14
5.2.2 Syntax.....	14
5.2.3 Semantics.....	14
5.3 Extended type box.....	15
5.3.1 Definition.....	15
5.3.2 Syntax.....	15
5.3.3 Semantics.....	15
6 Common specifications for movie files, item files and segment files	16
6.1 Data offsets.....	16
6.2 Box order.....	17
7 Movie file organization	21
7.1 Files, segments, and streams.....	21
7.2 Presentation structure.....	22
7.2.1 Box structure of a presentation.....	22
7.2.2 Meta data and media data.....	22
7.3 Structure-data.....	22
7.3.1 Box.....	22
7.3.2 Interpretation of matrix values.....	22
7.3.3 URIs as type indicators.....	23
7.3.4 Box order.....	23
7.4 Time structure overview.....	24
7.5 Identifiers.....	24
7.6 Uniform resource locators (URLs).....	24
8 Box structures	25
8.1 File structure and general boxes.....	25
8.1.1 Media data box.....	25
8.1.2 Free space box.....	25
8.1.3 Progressive download information box.....	26
8.1.4 Identified media data box.....	26
8.2 Movie structure.....	27
8.2.1 Movie box.....	27
8.2.2 Movie header box.....	27
8.3 Track structure.....	29
8.3.1 Track box.....	29
8.3.2 Track header box.....	29

8.3.3	Track reference box	32
8.3.4	Track group box	34
8.3.5	Track type box	36
8.4	Track media structure	37
8.4.1	Media box	37
8.4.2	Media header box	37
8.4.3	Handler box	38
8.4.4	Media information box	39
8.4.5	Media information header boxes	39
8.4.6	Extended language tag	40
8.5	Sample tables	40
8.5.1	Sample table box	40
8.5.2	Sample description box	41
8.5.3	Degradation priority box	43
8.5.4	Sample scale box	43
8.6	Track time structures	43
8.6.1	Time to sample boxes	43
8.6.2	Sync sample box	49
8.6.3	Shadow sync	50
8.6.4	Independent and disposable samples box	52
8.6.5	Edit box	53
8.6.6	Edit list box	54
8.7	Track data layout structures	56
8.7.1	Data information box	56
8.7.2	Data reference box	57
8.7.3	Sample size boxes	59
8.7.4	Sample to chunk box	60
8.7.5	Chunk offset box	61
8.7.6	Padding bits box	62
8.7.7	Sub-sample information box	62
8.7.8	Sample auxiliary information sizes box	65
8.7.9	Sample auxiliary information offsets box	67
8.7.10	Sub-sample Reference Table Box	68
8.8	Movie fragments	69
8.8.1	Movie extends box	69
8.8.2	Movie extends header box	70
8.8.3	Track extends box	71
8.8.4	Movie fragment box	72
8.8.5	Movie fragment header box	72
8.8.6	Track fragment box	73
8.8.7	Track fragment header box	73
8.8.8	Track fragment run box	75
8.8.9	Movie fragment random access box	76
8.8.10	Track fragment random access box	77
8.8.11	Movie fragment random access offset box	78
8.8.12	Track fragment decode time box	79
8.8.13	Level assignment box	80
8.8.14	Sample auxiliary information in movie fragments	81
8.8.15	Track extension properties box	82
8.8.16	Alternative startup sequence properties box	82
8.8.17	Metadata and user data in movie fragments	83
8.8.18	Redundant Sample Original Timing	83
8.9	Sample group structures	85
8.9.1	Overview	85
8.9.2	Sample to group box	85
8.9.3	Sample group description box	87
8.9.4	Representation of group structures in movie fragments	89
8.9.5	Compact sample to group box	90
8.10	User data	92

8.10.1	User data box	92
8.10.2	Copyright box	93
8.10.3	Track selection box	93
8.10.4	Track kind	95
8.10.5	Label box	95
8.11	Metadata support	96
8.11.1	MetaBox	96
8.11.2	XML boxes	98
8.11.3	Item location box	98
8.11.4	Primary item box	102
8.11.5	Item protection box	102
8.11.6	Item information box	103
8.11.7	Additional metadata container box	105
8.11.8	Metabox Relation box	105
8.11.9	URL forms for MetaBoxes	105
8.11.10	Static metadata	106
8.11.11	Item data box	107
8.11.12	Item reference box	107
8.11.13	Auxiliary video metadata	108
8.11.14	Item properties box	108
8.11.15	Brand item property	110
8.11.16	Handler property	110
8.11.17	T.35 item	111
8.12	File delivery format support	111
8.12.1	Overview	111
8.12.2	FD item information box	112
8.12.3	File partition box	113
8.12.4	FEC reservoir box	114
8.12.5	FD session group box	115
8.12.6	Group ID to name box	116
8.12.7	File reservoir box	116
8.13	Sub tracks	117
8.13.1	Overview	117
8.13.2	Backward compatibility	117
8.13.3	Sub track box	118
8.13.4	Sub track information box	118
8.13.5	Sub track definition box	119
8.13.6	Sub track sample group box	120
8.14	Segments	120
8.14.1	Overview	120
8.14.2	Segment type box	120
8.14.3	Segment index box	121
8.14.4	Subsegment index box	124
8.14.5	Producer reference time box	126
8.15	Entity grouping	127
8.15.1	General	127
8.15.2	Groups list box	128
8.15.3	Entity to group box	128
8.15.4	Entity group definitions	129
8.16	Compressed boxes	131
8.16.1	Overview and processing	131
8.16.2	Processing model	131
8.16.3	General syntax	133
8.16.4	General semantics	133
8.16.5	Original file-type box	133
8.16.6	Compressed movie box	134
8.16.7	Compressed movie fragment box	134
8.16.8	Compressed segment index box	134
8.16.9	Compressed subsegment index box	135

9	Hint track formats	135
9.1	RTP and SRTP hint track format	135
9.1.1	Overview	135
9.1.2	Sample entry format	136
9.1.3	Sample format	138
9.1.4	SDP information	140
9.1.5	Statistical information	141
9.2	ALC/LCT and FLUTE hint track format	141
9.2.1	Overview	141
9.2.2	Design principles	142
9.2.3	Sample entry format	143
9.2.4	Sample format	144
9.3	MPEG-2 transport hint track format	147
9.3.1	Overview	147
9.3.2	Design principles	147
9.3.3	Sample entry format	149
9.3.4	Sample format	151
9.3.5	Protected MPEG 2 transport stream hint track	153
9.4	RTP, RTCP, SRTP and SRTCP reception hint tracks	153
9.4.1	RTP reception hint track	153
9.4.2	RTCP reception hint track	156
9.4.3	SRTP reception hint track	158
9.4.4	SRTCP reception hint tracks	159
9.4.5	Protected RTP reception hint track	160
9.4.6	Recording procedure	160
9.4.7	Parsing procedure	161
10	Sample groups	161
10.1	Random access recovery points	161
10.1.1	Definition	161
10.1.2	Syntax	161
10.1.3	Semantics	161
10.2	Rate share groups	162
10.2.1	Overview	162
10.2.2	Rate share sample group description entry	163
10.2.3	Relationship between tracks	164
10.2.4	Bitrate allocation	164
10.3	Alternative startup sequences	165
10.3.1	Definition	165
10.3.2	Syntax	165
10.3.3	Semantics	166
10.3.4	Examples	166
10.4	Random access point (RAP) sample group	168
10.4.1	Definition	168
10.4.2	Syntax	169
10.4.3	Semantics	169
10.5	Temporal level sample group	169
10.5.1	Definition	169
10.5.2	Syntax	169
10.5.3	Semantics	169
10.6	Stream access point sample group	169
10.6.1	Definition	169
10.6.2	Syntax	170
10.6.3	Semantics	170
10.7	Sample-to-item sample group	170
10.7.1	Definition	170
10.7.2	Syntax	170
10.7.3	Semantics	170
10.8	Dependent random access point (DRAP) sample group	171

10.8.1	Definition	171
10.8.2	Syntax	171
10.8.3	Semantics	171
10.9	Pixel Aspect Ratio Sample Grouping	172
10.9.1	Definition	172
10.9.2	Syntax	172
10.9.3	Semantics	172
10.10	Clean Aperture Sample Grouping	172
10.10.1	Definition	172
10.10.2	Syntax	172
10.10.3	Semantics	172
10.11	EDRAP sample group	173
10.11.1	Definition	173
10.11.2	Syntax	173
10.11.3	Semantics	173
10.12	Essential descriptions hierarchy sample grouping	174
10.12.1	Definition	174
10.12.2	Syntax	175
10.12.3	Semantics	175
10.13	T.35 sample group	175
10.13.1	Definition	175
10.13.2	Syntax	176
10.13.3	Semantics	176
11	Derived file formats	176
12	Media-specific definitions	177
12.1	Video media	177
12.1.1	Media handler	177
12.1.2	Video media header	177
12.1.3	Sample entry	178
12.1.4	Pixel aspect ratio and clean aperture	178
12.1.5	Colour information	180
12.1.6	Content light level	181
12.1.7	Mastering display colour volume	181
12.1.8	Content colour volume	181
12.1.9	Ambient viewing environment	182
12.2	Audio media	182
12.2.1	Media handler	182
12.2.2	Sound media header	183
12.2.3	Sample entry	183
12.2.4	Channel layout	185
12.2.5	Downmix instructions	187
12.2.6	DRC information	191
12.2.7	Audio stream loudness	191
12.2.8	Audio rendering indication box	194
12.2.9	Audio element box	194
12.2.10	Audio element description box	195
12.2.11	Audio element positioning interactivity polar box	195
12.2.12	Audio element prominence interactivity box	197
12.2.13	Audio element selection box	197
12.2.14	Audio element selection description box	198
12.3	Metadata media	198
12.3.1	Media handler	198
12.3.2	Media header	198
12.3.3	Sample entry	198
12.4	Hint media	200
12.4.1	Overview	200
12.4.2	Media handler	201
12.4.3	Hint media header	201

12.4.4	Sample entry	202
12.5	Text media	202
12.5.1	Media handler	202
12.5.2	Media header	202
12.5.3	Sample entry	202
12.6	Subtitle media	203
12.6.1	Media handler	203
12.6.2	Subtitle media header	203
12.6.3	Sample entry	203
12.7	Font media	204
12.7.1	Media handler	204
12.7.2	Media header	204
12.7.3	Sample entry	204
12.8	Multiplexed timed metadata tracks	204
12.8.1	General	204
12.8.2	Overall design	205
12.8.3	Sample format	205
12.8.4	Sample entry format	205
12.8.5	Defined formats	208
12.9	Volumetric visual media	209
12.9.1	Media handler	209
12.9.2	Media header	209
12.9.3	Sample entry	209
12.9.4	Sample format	210
12.10	Haptic media	210
12.10.1	Media handler	210
12.10.2	Media header	210
12.10.3	Sample entry	210
12.10.4	Sample format	210
13	Transformed media tracks	210
13.1	General	210
13.2	Multiple transformations for a single transformed media track	211
13.3	Determining the untransformed sample entry type	211
13.3.1	General	211
13.3.2	Example for protected and restricted media (informative)	211
13.3.3	The 'codecs' MIME parameter for a transformed media track	212
13.4	Support for protected streams	212
13.4.1	Overview	212
13.4.2	Protection scheme information box	213
13.4.3	Original format box	214
13.4.4	IPMPInfoBox	214
13.4.5	IPMP control box	214
13.4.6	Scheme type box	214
13.4.7	Scheme information box	215
13.4.8	Scramble Scheme Information Box	215
13.5	Restricted media tracks	216
13.5.1	General	216
13.5.2	Restricted sample entry transformation	216
13.5.3	Restricted scheme information box	217
13.5.4	Scheme for stereoscopic video arrangements	217
13.5.5	Compatible scheme type box	220
13.5.6	Sample-packed tracks	220
13.6	Support for incomplete tracks	221
13.6.1	General	221
13.6.2	Transformation	222
13.6.3	Complete track information box	222
Annex A (informative) Overview and tutorial		223

Annex B (informative) Guidance on deriving from this document	231
Annex C (normative) Fragment identifiers for ISO base media resources	240
Annex D (informative) Management of extension code-points	241
Annex E (normative) File format brands	242
Annex F (normative) MIME type registration of segments	255
Annex G (informative) URI-labelled metadata forms	256
Annex H (informative) Processing of RTP streams and reception hint tracks	258
Annex I (normative) Stream access points	273
Annex J (informative) Segment index examples	276
Annex K (normative) Use of IETF RFC 6381 for ISOBMFF files	281
Bibliography	284

Sample Document

get full document from standards.iteh.ai

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 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 eighth edition cancels and replaces the seventh edition (ISO/IEC 14496-12:2022), which has been technically revised.

The main changes are as follows:

- specification of essential sample groups
- addition of the essential descriptions hierarchy sample group
- addition of the preselection entity group and related boxes
- addition of the extended dependent random access point (EDRAP) sample group
- specification of the sample-packed restricted video track
- addition of the associated external stream track reference
- relaxing the presence requirements of the `HandlerBox` within a `MetaBox`
- addition of the handler property for items.

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.

Introduction

The ISO base media file format is designed to contain timed media information for a presentation in a flexible, extensible format that facilitates interchange, management, editing, and presentation of the media. This presentation may be 'local' to the system containing the presentation, or may be via a network or other stream delivery mechanism.

The file structure is object-oriented; a file can be decomposed into constituent objects very simply, and the structure of the objects inferred directly from their type.

The file format is designed to be independent of any particular network protocol while enabling efficient support for them in general.

The ISO base media file format is a base format for media file formats.

Structure of this document

[Clause 4](#) defines ISO base media files; files that are built from boxes.

[Clause 5](#) defines typed ISO base media files, i.e. ISO base media files with a `FileTypeBox`.

[Clause 6](#) provides common specifications to movie files, item files and segment files.

[Clause 7](#) gives the core concepts and data-types for time-based presentations, called 'movies' in this document.

[Clause 8](#) defines the boxes used by time-based presentations, and other formats.

[Clause 9](#) defines the hint track formats used to support some streaming protocols.

[Clause 10](#) builds on the concept of sample groups as defined in [subclause 8.9](#) and defines some sample groups.

[Clause 11](#) defines how to base a file format on this document.

[Clause 12](#) builds on the general concepts of tracks as defined in [Clause 8](#), and defines track formats for various general types of media (video, sound, etc.).

[Clause 13](#) specifies different types of transformed media tracks.

[Annex A](#) provides an informative introduction to time-based presentations, which may be of assistance to first-time readers and implementers.

[Annex B](#) provides guidance on writing derived specifications.

[Annex C](#) provides the syntax for uniform resource identifier (URI) fragments.

[Annex D](#) documents how identifier values defined externally to this document are managed.

[Annex E](#) defines brands that may be used to identify the conformance and reader requirements to the structures defined in this document for time-based presentations.

[Annex F](#) contains the formal IANA registration of segments.

[Annex G](#) defines some forms used for labelling metadata with uniform resource identifier (URI) labels.

[Annex H](#) provides an overview of the use of hint tracks for RTP streams and RTP stream reception.

[Annex I](#) contains the formal definitions of the types of stream access points in timed media streams.

[Annex J](#) contains examples of the use of the `SegmentIndexBox` defined in [8.14.3](#).

[Annex K](#) defines the MIME parameters that may be used to annotate MIME types for time-based presentations based on [Clause 6](#).

Sample Document

get full document from standards.iteh.ai

Information technology — Coding of audio-visual objects —

Part 12: ISO base media file format

1 Scope

This document specifies the ISO base media file format, which is a general format forming the basis for a number of other more specific file formats. This format contains the timing, structure, and media information for timed sequences of media data, such as audio-visual presentations.

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, *Code for individual languages and language groups*

ITU-T X.667 | ISO/IEC 9834-8, *Information technology — Procedures for the operation of object identifier registration authorities — Part 8: Generation of universally unique identifiers (UUIDs) and their use in object identifiers*

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

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

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

ISO 15076-1, *Image technology colour management — Architecture, profile format and data structure — Part 1: Based on ICC.1:2010*

ISO/IEC 15938-1, *Information technology — Multimedia content description interface — Part 1: Systems*

ISO/IEC 23001-1, *Information technology — MPEG systems technologies — Part 1: Binary MPEG format for XML*

ISO/IEC 23001-14, *Information technology — MPEG systems technologies — Part 14: Partial file format*

ISO/IEC 23002-3, *Information technology — MPEG video technologies — Part 3: Representation of auxiliary video and supplemental information*

ISO/IEC 23003-4, *Information technology — MPEG audio technologies — Part 4: Dynamic range control*

ITU-T H.265 | ISO/IEC 23008-2, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

ISO/IEC 23091-2, *Information technology — Coding-independent code points — Part 2: Video*

ISO/IEC 23091-3, *Information technology — Coding-independent code points — Part 3: Audio*

ITU-R TF.460-6:2002, *Standard-frequency and time-signal emissions*

ITU-R BS.1770-4, *Algorithms to measure audio programme loudness and true-peak audio level*

ITU-T T.35, *Procedure for the allocation of ITU-T defined codes for non-standard facilities*

IETF BCP 47, *Tags for Identifying Languages*

IETF RFC 1951, *DEFLATE Compressed Data Format Specification version 1.3*

IETF RFC 2045, *Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies*

IETF RFC 2046, *Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types*

IETF RFC 3061, *A URN Namespace of Object Identifiers*

IETF RFC 3629, *UTF-8, a transformation format of ISO 10646*

IETF RFC 3711:2004, *The Secure Real-time Transport Protocol (SRTP)*

IETF RFC 4122, *A Universally Unique Identifier (UUID) URN Namespace*

IETF RFC 5052, *Forward Error Correction (FEC) Building Block*

IETF RFC 5905, *Network Time Protocol Version 4: Protocol and Algorithms Specification*

W3C Recommendation, *Extensible Markup Language (XML) 1.0 (Fifth Edition)*, 26 November 2008, <https://www.w3.org/TR/2008/REC-xml-20081126/>

3 Terms, definitions and abbreviated terms

3.1 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/>

3.1.1

access unit

media data pertaining to a particular composition time in a media stream, usually carried in one sample of a media track

3.1.2

bitstream

<SAP information> file or segment compatible with the ISO base media file format

Note 1 to entry: This definition applies for the interpretation of SAP information for files and segments compatible with the ISO base media file format.

3.1.3

box

object-oriented building block defined by a unique type identifier and length

Note 1 to entry: Called 'atom' in some specifications, including the first definition of MP4.

3.1.4

child box

box contained in another box

3.1.5

chunk

contiguous set of samples for one track

3.1.6

clean aperture

part of a decoded video image from which undesirable pixels introduced for coding purposes such as having integer number of coding blocks have been removed for presentation

3.1.7

composition timestamp

timestamp on the media timeline such that samples are presented in composition timestamp order and that establishes their relative composition timing

3.1.8

container

<SAP information> file or segment compatible with the ISO base media file format

Note 1 to entry: This definition applies for the interpretation of SAP information for files and segments compatible with the ISO base media file format.

3.1.9

container box

box whose sole purpose is to contain and group a set of related boxes

Note 1 to entry: Container boxes are normally not derived from `FullBox`

3.1.10

decoding timestamp

timestamp on the media timeline such that samples are decoded in decoding timestamp order

Note 1 to entry: The decoding timestamps primarily define the required decoding order; systems may decode at a time of their choosing; in the case of signed composition offsets, decoding timestamps do not necessarily precede composition timestamps

3.1.11

dual-headed file

typed ISO base media file that is both an item file and a movie file

3.1.12

EDRAP sample

sample for which all subsequent samples in both decoding and output order can be correctly decoded provided that the required preceding SAP or EDRAP samples are available

3.1.13

essential sample group

sample group that describes essential information for the associated samples and is associated with a sample group description for which the `version` field is equal to 3

3.1.14

file level

byte position in an ISO base media file not contained in a `Box` structure

3.1.15

full aperture

decoded video image as output by the decoder which may contain undesirable pixels for presentation

3.1.16

haptic media

timed tactile signals to be presented as part of the media presentation

3.1.17

hint track

special track which does not contain media data, but instead contains instructions for packaging one or more tracks into a streaming channel