ISO/IEC 14496-12:2022-01 (E)

Information technology - Coding of audio-visual objects - Part 12: ISO base media file format

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 7

4 Object-structured file organization **7**
4.1 File structure 7
4.2 Object structure 7
4.2.1 Object syntax conventions 7
4.2.2 Object definitions 8
4.2.3 Extensibility of object definitions 9
4.3 File-type box 10
4.3.1 Definition 10
4.3.2 Syntax 10
4.3.3 Semantics 11
4.4 Extended type box 11
4.4.1 Definition 11
4.4.2 Syntax 11
4.4.3 Semantics 11

5 Structure of this document **12**

6 ISO base media file organization **12**
6.1 Files, segments, and streams 12
6.2 Presentation structure 13
6.2.1 Object structure of a presentation 13
6.2.2 Meta data and media data 13
6.3 Structure-data (objects) 13
6.3.1 Box 13
6.3.2 Data types and fields 13
6.3.3 URIs as type indicators 14
6.3.4 Box order 15
6.4 Time structure overview 18
6.5 Identifiers 19
6.6 Brand identification 19
6.7 Uniform resource locators (URLs) 19

7 Streaming support **19**

8 Box structures **19**
8.1 File structure and general boxes 19
8.1.1 Media data box 19
8.1.2 Free space box 20
8.1.3 Progressive download information box 20
8.1.4 Identified media data box 21
8.2 Movie structure 21
8.2.1 Movie box 21
8.2.2 Movie header box 21

8.3 Track structure 23
8.3.1 Track box 23
8.3.2 Track header box 23
8.3.3 Track reference box 26
8.3.4 Track group box 28
8.3.5 Track type box 29
8.4 Track media structure 30
8.4.1 Media box 30
8.4.2 Media header box 30
8.4.3 Handler reference box 31
8.4.4 Media information box 32
8.4.5 Media information header boxes 32
8.4.6 Extended language tag 32
8.5 Sample tables 33
8.5.1 Sample table box 33
8.5.2 Sample description box 34
8.5.3 Degradation priority box 36
8.5.4 Sample scale box 36
8.6 Track time structures 36
8.6.1 Time to sample boxes 36
8.6.2 Sync sample box 41
8.6.3 Shadow sync 42
8.6.4 Independent and disposable samples box 43
8.6.5 Edit box 45
8.6.6 Edit list box 45
8.7 Track data layout structures 48
8.7.1 Data information box 48
8.7.2 Data reference box 48
8.7.3 Sample size boxes 50
8.7.4 Sample to chunk box 51
8.7.5 Chunk offset box 52
8.7.6 Padding bits box 52
8.7.7 Sub-sample information box 53
8.7.8 Sample auxiliary information sizes box 54
8.7.9 Sample auxiliary information offsets box 56
8.8 Movie fragments 57
8.8.1 Movie extends box 57
8.8.2 Movie extends header box 58
8.8.3 Track extends box 58
8.8.4 Movie fragment box 59
8.8.5 Movie fragment header box 60
8.8.6 Track fragment box 60
8.8.7 Track fragment header box 60
8.8.8 Track fragment run box 62
8.8.9 Movie fragment random access box 63
8.8.10 Track fragment random access box 64
8.8.11 Movie fragment random access offset box 65
8.8.12 Track fragment decode time box 65
8.8.13 Level assignment box 66
8.8.14 Sample auxiliary information in movie fragments 68
8.8.15 Track Extension Properties box 68
8.8.16 Alternative startup sequence properties box 68
8.8.17 Metadata and user data in movie fragments 69
8.9 Sample group structures 70
8.9.1 Overview 70
8.9.2 Sample to group box 70
8.9.3 Sample group description box 72
8.9.4 Representation of group structures in movie fragments 74
8.9.5 Compact sample to group box 75
8.10 User data 77
8.10.1 User data box 77
8.10.2 Copyright box 77
8.10.3 Track selection box 78

8.10.4 Track kind 79
8.11 Metadata support 80
8.11.1 MetaBox 80
8.11.2 XML boxes 81
8.11.3 Item location box 81
8.11.4 Primary item box 84
8.11.5 Item protection box 85
8.11.6 Item information box 85
8.11.7 Additional metadata container box 87
8.11.8 Metabox Relation box 87
8.11.9 URL forms for MetaBoxes 88
8.11.10 Static metadata 88
8.11.11 Item data box 89
8.11.12 Item reference box 89
8.11.13 Auxiliary video metadata 90
8.11.14 Item properties box 90
8.11.15 Brand item property 92
8.12 Support for protected streams 93
8.12.1 Overview 93
8.12.2 Protection scheme information box 94
8.12.3 Original format box 94
8.12.4 IPMPInfoBox 95
8.12.5 IPMP control box 95
8.12.6 Scheme type box 95
8.12.7 Scheme information box 95
8.12.8 Scramble Scheme Information Box 96
8.13 File delivery format support 96
8.13.1 Overview 96
8.13.2 FD item information box 97
8.13.3 File partition box 97
8.13.4 FEC reservoir box 99
8.13.5 FD session group box 99
8.13.6 Group ID to name box 100
8.13.7 File reservoir box 101
8.14 Sub tracks 101
8.14.1 Overview 101
8.14.2 Backward compatibility 102
8.14.3 Sub track box 102
8.14.4 Sub track information box 102
8.14.5 Sub track definition box 103
8.14.6 Sub track sample group box 104
8.15 Post-decoder requirements on media 104
8.15.1 General 104
8.15.2 Restricted sample entry transformation 105
8.15.3 Restricted scheme information box 105
8.15.4 Scheme for stereoscopic video arrangements 106
8.15.5 Compatible scheme type box 108
8.16 Segments 108
8.16.1 Overview 108
8.16.2 Segment type box 108
8.16.3 Segment index box 109
8.16.4 Subsegment index box 112
8.16.5 Producer reference time box 114
8.17 Support for incomplete tracks 115
8.17.1 General 115
8.17.2 Transformation 116
8.17.3 Complete track information box 116
8.18 Entity grouping 117

8.18.1 General 117
8.18.2 Groups list box 117
8.18.3 Entity to group box 117
8.19 Compressed boxes 118
8.19.1 Overview and processing 118
8.19.2 Processing model 119
8.19.3 General syntax 120
8.19.4 General semantics 120
8.19.5 Original file-type box 120
8.19.6 Compressed movie box 121
8.19.7 Compressed movie fragment box 121
8.19.8 Compressed segment index box 121
8.19.9 Compressed subsegment index box 122

9 Hint track formats 122
9.1 RTP and SRTP hint track format 122
9.1.1 Overview 122
9.1.2 Sample description format 123
9.1.3 Sample format 124
9.1.4 SDP information 127
9.1.5 Statistical information 127
9.2 ALC/LCT and FLUTE hint track format 128
9.2.1 Overview 128
9.2.2 Design principles 129
9.2.3 Sample description format 130
9.2.4 Sample format 130
9.3 MPEG-2 transport hint track format 133
9.3.1 Overview 133
9.3.2 Design principles 134
9.3.3 Sample description format 135
9.3.4 Sample format 137
9.3.5 Protected MPEG 2 transport stream hint track 139
9.4 RTP, RTCP, SRTP and SRTCP reception hint tracks 140
9.4.1 RTP reception hint track 140
9.4.2 RTCP reception hint track 143
9.4.3 SRTP reception hint track 144
9.4.4 SRTCP reception hint tracks 146
9.4.5 Protected RTP reception hint track 147
9.4.6 Recording procedure 147
9.4.7 Parsing procedure 147

10 Sample groups 147
10.1 Random access recovery points 147
10.1.1 Definition 147
10.1.2 Syntax 148
10.1.3 Semantics 148
10.2 Rate share groups 148
10.2.1 Overview 148
10.2.2 Rate share sample group entry 149
10.2.3 Relationship between tracks 150
10.2.4 Bitrate allocation 151
10.3 Alternative startup sequences 151
10.3.1 Definition 151
10.3.2 Syntax 152
10.3.3 Semantics 152
10.3.4 Examples 152
10.4 Random access point (RAP) sample group 154
10.4.1 Definition 154
10.4.2 Syntax 154

10.4.3 Semantics 154
10.5 Temporal level sample group 154
10.5.1 Definition 154
10.5.2 Syntax 155
10.5.3 Semantics 155
10.6 Stream access point sample group 155
10.6.1 Definition 155
10.6.2 Syntax 155
10.6.3 Semantics 155
10.7 Sample-to-item sample group 156
10.7.1 Definition 156
10.7.2 Syntax 156
10.7.3 Semantics 156
10.8 Dependent random access point (DRAP) sample group 156
10.8.1 Definition 156
10.8.2 Syntax 157
10.8.3 Semantics 157
10.9 Pixel Aspect Ratio Sample Grouping 157
10.9.1 Definition 157
10.9.2 Syntax 157
10.9.3 Semantics 157
10.10 Clean Aperture Sample Grouping 157
10.10.1 Definition 157
10.10.2 Syntax 158
10.10.3 Semantics 158

11 Derived file formats 158

12 Media-specific definitions 159
12.1 Video media 159
12.1.1 Media handler 159
12.1.2 Video media header 159
12.1.3 Sample entry 159
12.1.4 Pixel aspect ratio and clean aperture 160
12.1.5 Colour information 162
12.1.6 Content light level 163
12.1.7 Mastering display colour volume 163
12.1.8 Content colour volume 163
12.1.9 Ambient viewing environment 164
12.2 Audio media 164
12.2.1 Media handler 164
12.2.2 Sound media header 164
12.2.3 Sample entry 165
12.2.4 Channel layout 167
12.2.5 Downmix instructions 169
12.2.6 DRC information 172
12.2.7 Audio stream loudness 173
12.3 Metadata media 175
12.3.1 Media handler 175
12.3.2 Media header 175
12.3.3 Sample entry 175
12.4 Hint media 177
12.4.1 Overview 177
12.4.2 Media handler 178
12.4.3 Hint media header 178
12.4.4 Sample entry 178
12.5 Text media 179
12.5.1 Media handler 179
12.5.2 Media header 179

12.5.3 Sample entry 179
12.6 Subtitle media 179
12.6.1 Media handler 179
12.6.2 Subtitle media header 179
12.6.3 Sample entry 180
12.7 Font media 181
12.7.1 Media handler 181
12.7.2 Media header 181
12.7.3 Sample entry 181
12.8 Transformed media 181
12.8.1 General 181
12.8.2 Multiple transformations for a single transformed media track 181
12.8.3 Determining the untransformed sample entry type 182
12.8.4 The `'codecs'` MIME parameter for a transformed media track 182
12.9 Multiplexed timed metadata tracks 182
12.9.1 General 182
12.9.2 Overall design 182
12.9.3 Sample format 183
12.9.4 Sample entry format 183
12.9.5 Defined formats 186
12.10 Volumetric visual media 187
12.10.1 Media handler 187
12.10.2 Media header 187
12.10.3 Sample entry 187
12.10.4 Sample format 188
12.11 Haptic media 188
12.11.1 Media handler 188
12.11.2 Media header 188
12.11.3 Sample entry 188
12.11.4 Sample format 188

Annex A (informative) **Background and tutorial** **189**

Annex B (informative) **Guidance on deriving from this document** **197**

Annex C (normative) **Fragment identifiers for ISO base media resources** **206**

Annex D (informative) **Management of extension code-points** **207**

Annex E (normative) **File format brands** **209**

Annex F (normative) **MIME type registration of segments** **220**

Annex G (informative) **URI-labelled metadata forms** **221**

Annex H (informative) **Processing of RTP streams and reception hint tracks** **223**

Annex I (normative) **Stream access points** **240**

Annex J (informative) **Segment index examples** **243**

Annex K (normative) **Use of IETF RFC 6381 for ISOBMFF files** **246**

Bibliography **249**