INTERNATIONAL STANDARD

ISO/IEC 23000-2

Second edition
2008-01-15

Corrected version
2008-04-15

Information technology — Multimedia application format (MPEG-A) —

Part 2: MPEG music player application format

Technologies de l'information — Format pour application multimédia (MPEG-A) —
Partie 2: Format pour application musicienne MPEG

Reference number
ISO/IEC 23000-2:2008(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 23000-2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC 23000-2:2006), which has been technically revised.

ISO/IEC 23000 consists of the following parts, under the general title *Information technology — Multimedia application format (MPEG-A)*:

— *Part 1: Purpose for multimedia application formats* [Technical Report]

— *Part 2: MPEG music player application format*

— *Part 3: MPEG photo player application format*

— *Part 4: Musical slide show application format*

— *Part 5: Media streaming application format*

— *Part 7: Open access application format*

— *Part 8: Portable video application format*

— *Part 9: Digital multimedia broadcasting application format*

— *Part 10: Video surveillance application format*

This corrected version of ISO/IEC 23000-2:2008 incorporates the following corrections: the provision of a CD-ROM containing reference software for multimedia application format; an updated list of parts in the Foreword.

Introduction

MPEG has developed a number of standards, all of which strive to serve the needs of consumers and industry. Among those are MPEG-4, a next-generation suite of standards for media compression, and MPEG-7, a suite of standards for meta-data representation. MPEG-4 specifies what MPEG expects to be another very successful specification, the MPEG-4 File Format, while MPEG-7 specifies not only signal-derived meta-data, but also archival meta-data such as Artist, Album and Song Title.

As such, MPEG-4 and MPEG-7 represent an ideal environment to support the current “MP3 music library” user experience, and, moreover, to extend that experience in new directions.

Firstly, this part of ISO/IEC 23000 shows how to carry MP3 information (music and meta-data) within the MPEG-4 and MPEG-7 framework. Moving MP3 into the MPEG-4 world supports, as a baseline, everything that users know and expect, but offers the capability to deliver a much richer music experience with components of MPEG-4, MPEG-7 and MPEG-21 at our disposal.

Secondly, this part of ISO/IEC 23000 builds on the music player and extends it to the Protected Music Player for both mp4 and mp21 file types including (1) mp4 file as protected content format with fixed encryption, without key management components and (2) protected mp21 file with flexible tool selection and key management components.

This part of ISO/IEC 23000 contains conformance and reference software.

Information technology — Multimedia application format (MPEG-A) —

Part 2:
MPEG music player application format

1 Scope

This part of ISO/IEC 23000 presents a basic architecture for constructing an annotated music library. It defines a simple file format for songs and a file format for albums and playlists. A conformant player application has to support all these specified file formats.

2 Normative references

The following referenced documents are indispensable for the application 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/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 13818-3:1998, *Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio*

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

ISO/IEC 14496-3:2005, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-12:2005, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-13:2004, *Information technology — Coding of audio-visual objects — Part 13: Intellectual Property Management and Protection (IPMP) extensions*

ISO/IEC 14496-14:2003, *Information technology — Coding of audio-visual objects — Part 14: MP4 file format*

ISO/IEC 15938-5:2003, *Information technology — Multimedia content description interface — Part 5: Multimedia description schemes*

ISO/IEC TR 15938-8:2002, *Information technology — Multimedia content description interface — Part 8: Extraction and use of MPEG-7 descriptions*

ISO/IEC 21000-2:2005, *Information technology — Multimedia framework (MPEG-21) — Part 2: Digital Item Declaration*

ISO/IEC 21000-4:2006, *Information technology — Multimedia framework (MPEG-21) — Part 4: Intellectual Property Management and Protection Components*

ISO/IEC 21000-5:2004, *Information technology — Multimedia framework (MPEG-21) — Part 5: Rights Expression Language*

ISO/IEC 21000-9:2005, *Information technology — Multimedia framework (MPEG-21) — Part 9: File Format*

3 Overview of MPEG Standards

3.1 MPEG-1 Layer III

ISO/IEC 11172-3:1993 specifies MPEG-1 Audio. From that specification, MPEG-1 Layer III (or MP3) is one of the most widely deployed MPEG audio standards ever. Its wide appeal is due to both its good compression performance and its simplicity of implementation. The vast majority of compressed music archives use MP3 encoding.

One aspect of the simplicity of Layer III is that it specifies a self-synchronizing transport, making it amenable to both storage in a computer file and transmission over a channel without byte framing. In the context of transmission channels, Layer III can operate over a constant-rate isocronous link, and has constant-rate headers (as does Layer I and II). However Layer III is an instantaneously-variable-rate coder, which adapts to the constant-rate channel by using a “bit buffer” and “back pointers.” Each of the headers signals the start of another block of audio signal, however due to the Layer III syntax, the data associated with that next block of audio signal may be in a prior segment of the bitstream, pointed to by the back pointer (see Figure 1, specifically the curved arrow pointing to main_data_begin). We note that this is in contrast to the MPEG-4 view of data stream segmentation, in which one Access Unit contains all information necessary to decode one segment of audio.

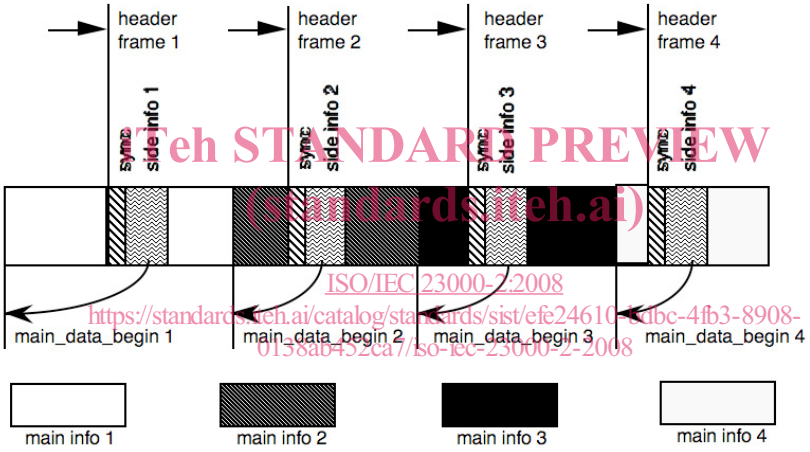

Figure 1 — Layer III bitstream organization

3.2 MPEG-4 “MPEG-1/2 Audio in MPEG-4”

ISO/IEC 14496-3:2005 subpart 9 “MPEG-1/2 Audio in MPEG-4” specifies a method for segmenting and formatting Layer III bitstreams into MPEG-4 Access Units, and therefore is often referred to as “MP3onMP4”. This consists primarily of re-arranging the compressed data associated with a given header such that it follows the header. This typically results in new segments that are no longer of constant length, but that is perfectly in accordance with the definition of MPEG-4 Access Units. See example in Figure 2.

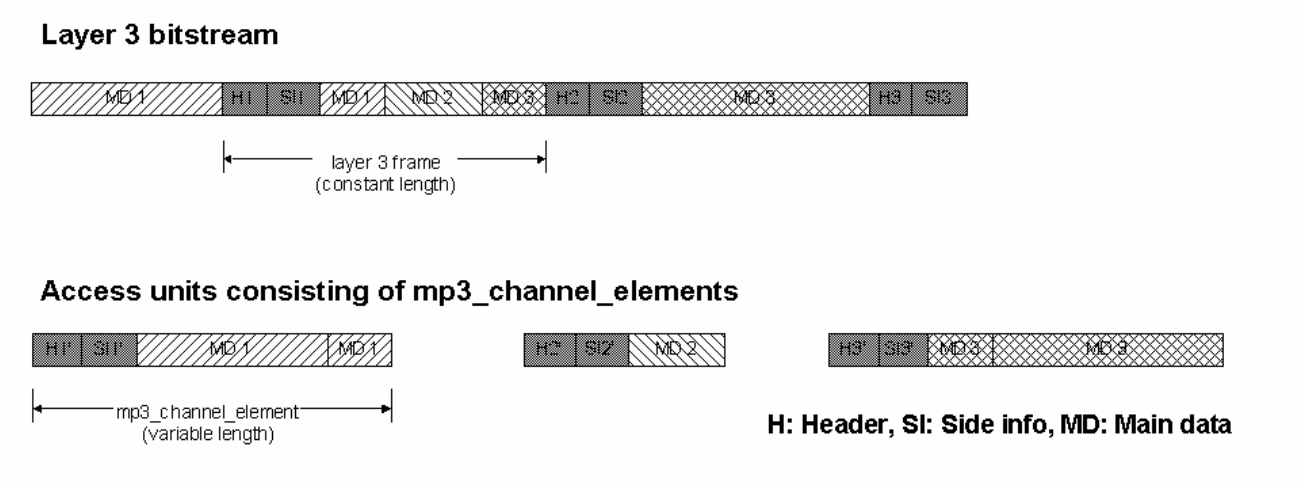

Figure 2 — Converting an MPEG-1/2 Layer 3 bitstream into mp3_channel_elements

3.3 ISO Base Media File Format

ISO/IEC 14496-12:2005 specifies the ISO Base Media File Format that 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. The ISO Base Media File Format is a base format for media file formats. In particular, the MPEG-4 file format derives from this base file format.

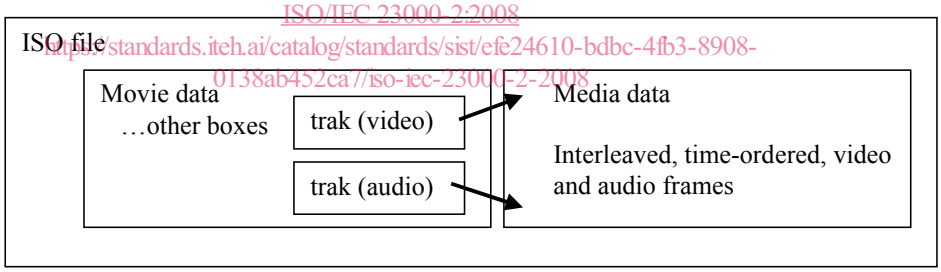

Figure 3 — Example of a simple ISO file used for interchange, containing two streams

The file structure is object-oriented as shown in Figure 3 which means that 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.

3.4 MPEG-4 File Format

ISO/IEC 14496-12:2005 and ISO/IEC 14496-14:2003 together specify the MPEG-4 File Format. This supports storage of compressed audio data (e.g. MP3onMP4) in tracks. It also provides support for metadata in the form of 'meta' boxes at the File, Movie and Track level. This allows support for static (un-timed) meta-data. Figure 4 schematically illustrates the location of these un-timed MPEG-7 Metadata boxes. 4.6 provides details as to when the Metadata boxes at each level are used.

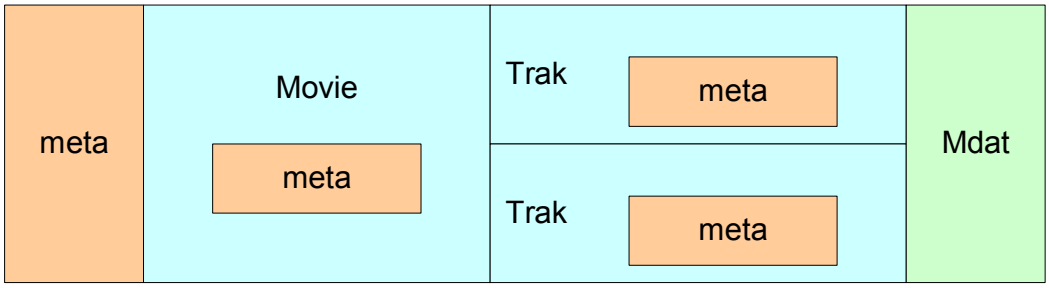

Figure 4 — Support of Static un-timed Metadata in ISO/MP4 Files

3.5 MPEG-21 File Format

ISO/IEC 21000-9:2005 specifies the MPEG-21 File Format. It uses the file-level 'meta' box of the ISO Base Media File Format to store an MPEG-21 Digital Item Declaration. The resources are either included in the file's mdat box or are externally referenced. The 'iloc' and the 'iinf' boxes in the meta box are used to address the resources. Figure 5 shows a simple example of an MPEG-21 file.

Figure 5 — Basic structure of an MPEG-21 File

3.6 MPEG-7 Multi-Media Description Scheme

ISO/IEC 15938-5:2003, the Multimedia Description Scheme (MDS) specifies all non-Visual and non-Audio specific metadata (e.g. Artist, Title, Date) in the MPEG-7 standard. As such it is able to represent all of the information found in the ID3v1 [3] metadata tagging format, as well as the corresponding information in the ID3v2 format, with its most popular version ID3v2.3 [4].

4 Song File Format

4.1 Introduction

This clause defines a format that contains a single music track with associated meta-data and a single still image containing cover art.

4.2 Audio track

The audio track contains encoded audio data according to ISO/IEC 11172-3 Layer III ("MP3") in the form of MPEG-4 audio samples.

ISO/IEC 11172-3 Layer III ("MP3") specifies a music compression scheme that results in a sequence of bits, forming a bitstream file. In contrast, ISO/IEC 14496-3:2005 specifies a music compression scheme that results in a sequence of packets (Access Units) that contain all data corresponding to one fraction of playout time. Access Units and their corresponding timing information is stored directly into the MPEG-4 File Format, specified in ISO/IEC 14496-14:2003. An MP3 bitstream can be translated into a series of MP3 Access Units using the "MP3onMP4" conversion as specified in ISO/IEC 14496-3:2005 subpart 9. This conversion allows to use mp3 encoded audio data like audio data encoded with any MPEG-4 audio codec type (e.g. AAC) within an MPEG-4 system, e.g. store the MP3 access units to an MPEG-4 file.

4.3 Meta-data

The MPEG-4 file format supports the storage of meta-data associated to a data track. Associated meta-data describing the audio track, like artist or song name, is expressed in MPEG-7 nomenclature, as specified in ISO/IEC 15938-5:2003. MP3 bitstream files can contain associated meta-data, typically ID3 tags [3,4]. The specific mapping from ID3 v1.1 tags and the corresponding ID3 v2.3 frames to MPEG-7 meta-data is show in Table 1. Parenthetical comments under Artist clarify that MPEG-7 is able to make a distinction between Artist as a *person* and Artist as a *group name*.

Table 1 — Mapping from ID3 v1.1 and ID3 v2.3 Tags to MPEG-7

ID3 v1	**ID3 v2.3 Frame**	**Description**	**MPEG-7 Path**
Artist	TOPE *(Original artist / performer)*	Artist performing the song	CreationInformation/Creation/Creator[Role/@href="urn:mpeg:mpeg7:RoleCS:2001:PERFORMER"]/Agent[@xsi:type="PersonType"]/Name/{FamilyName, GivenName} *(Artist Name)* CreationInformation/Creation/Creator[Role/@href="urn:mpeg:mpeg7:RoleCS:2001:PERFORMER"]/Agent[@xsi:type="PersonGroupType"]/Name *(Group Name)*
Album	TALB *(Album / Movie / Show title)*	Title of the album	CreationInformation/Creation/Title[@type="albumTitle"]
Song Title	TIT2 *(Title / Songname / Content description)*	Title of the song	CreationInformation/Creation/Title[@type="songTitle"]
Year	TORY (*Original release year)*	Year of the recording	CreationInformation/CreationCoordinates/Date/TimePoint (*Recording date.*)
Comment	COMM	Any comment of any length	CreationInformation/Creation/Abstract/FreeTextAnnotation
Track	TRCK *(Track number / Position in set)*	CD track number of song	Semantics/SemanticBase[@xsi:type="SemanticStateType"]/AttributeValuePair
Genre	TCON *(Content type)*	ID 3 V1.1 Genre ID 3 V2 Genre (4)(Eurodisco)	CreationInformation/Classification/Genre[@href="urn:id3:v1:4"] CreationInformation/Classification/Genre[@href="urn:id3:v1:4"]/Term[@termID="urn:id3:v2:Eurodisco"] CreationInformation/Classification/Genre[@href="urn:id3:v1:4"] CreationInformation/Classification/Genre[@type="secondary"][@href="urn:id3:v2:Eurodisco"]

MPEG-7 Path notation is a shorthand for the full XML notation, and an example of the correspondence between MPEG-7 Path and XML notation is shown in Annex A.

4.4 File Conversion

This subclause describes a lossless, reversible conversion of a standard mp3 bitstream file into an MPEG-4 file according to this Music Player Application Format. An MP3 bitstream file (containing mp3 audio frames and an ID3 tag) is converted using two modules, as shown in Figure 6.

The first module translates an MP3 bitstream into a series of MP3 Access Units. This is accomplished by the MP3onMP4 formatter, specified in ISO/IEC 14496-3:2005 subpart 9. The Access Units are stored into one audio track of an MPEG-4 File.

The second module extracts the meta-data information from the input file's ID3 tag and expresses it as MPEG-7 descriptor according to 4.3. This MPEG-7 meta-data is stored - together with the optional JPEG image for cover art - into the corresponding meta-box of the audio track.

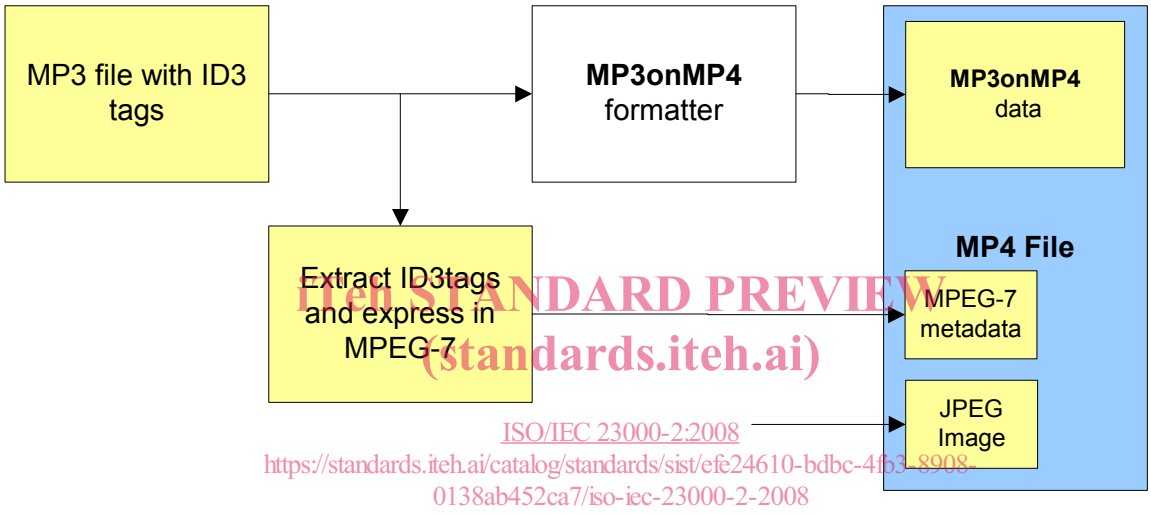

Figure 6 — Encoder System Architecture

4.5 Playback

Playback consists of

— extracting the meta-data from the MPEG-4 file and displaying it on a suitable visual interface.

— extracting the MP3onMP4 data from the MPEG-4 file, filtering it with very light-weight de-formatting operation, and playing it through a "classic" MP3 decoder.

In practice, it may be that the MP3onMP4 data is played by an "MP3onMP4 decoder," consisting of the concatenation of the MP3onMP4 deformatter and the MP3 decoder.

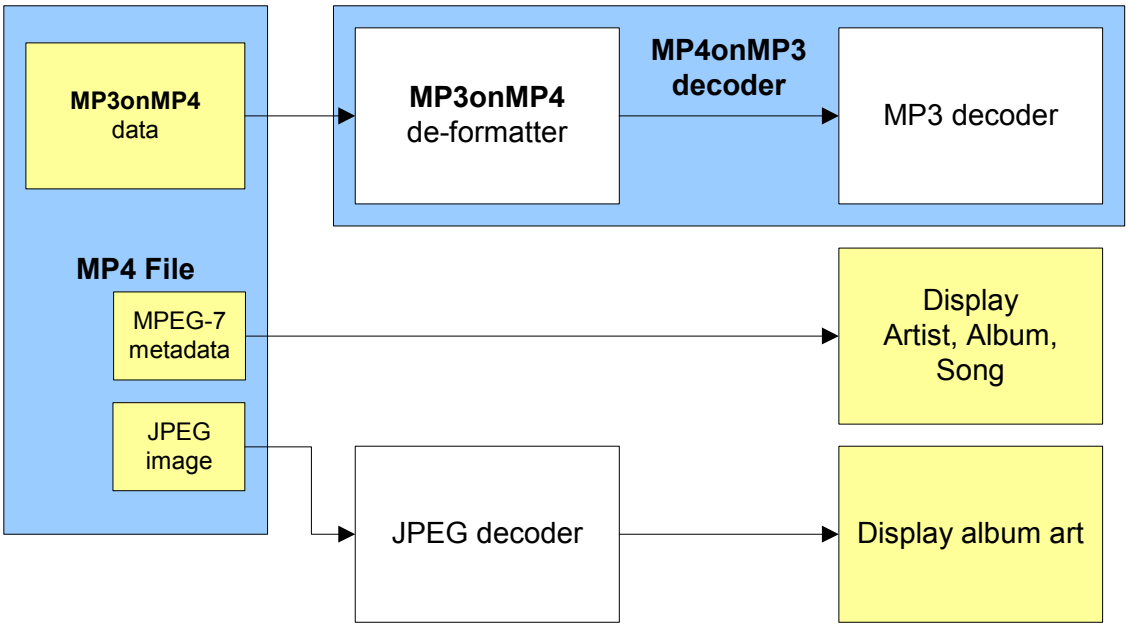

Figure 7 — Player System Architecture

4.6 File Structure

This subclause specifies the basic song file format of the Music Player Application Format in detail. The structure is based on the Playback model described above, however present a detailed view of the internals of the File Format (based on ISO/IEC 14496-12:2005), and highlight what resources a Playback application requires in order to decode the file.

For this structure (and the structures of the following Clause 5) the following information is given:

— File Example – a visual example of the file, showing the important boxes for playback (some boxes have been omitted to minimise complexity).

— File Type – A top level handler which indicates the type of file format that the structure uses. Hence if a Playback application supports either the major-brand and/or the compatible-brands *ftype* box field, it can decode this structure.

— Meta data Handler Type – MAF currently supports file level (*ftyp*) meta type handlers of mp7t and mp21.

— Resource Lookup/Playback – this is an additional section to the Playback model in order to highlight which ISO Base File Format boxes are used to decode the file.

— Notes – additional information that may aid understanding of the structure.

Additionally the following common notes apply to all structures:

NOTE The extent_count=1 for all items in the *iloc* box in each Resource Lookup/Playback section.

This subclause defines the basic structure of the Music Player Application Format specification: a MPEG-4 file with a single music track with optional JPEG image and MPEG-7 meta-data. The MPEG-7 Meta Data and optional JPEG Image are implicitly associated with the MP3 data by using the *trak* level *meta* box in the same track that contains the MP3 data.

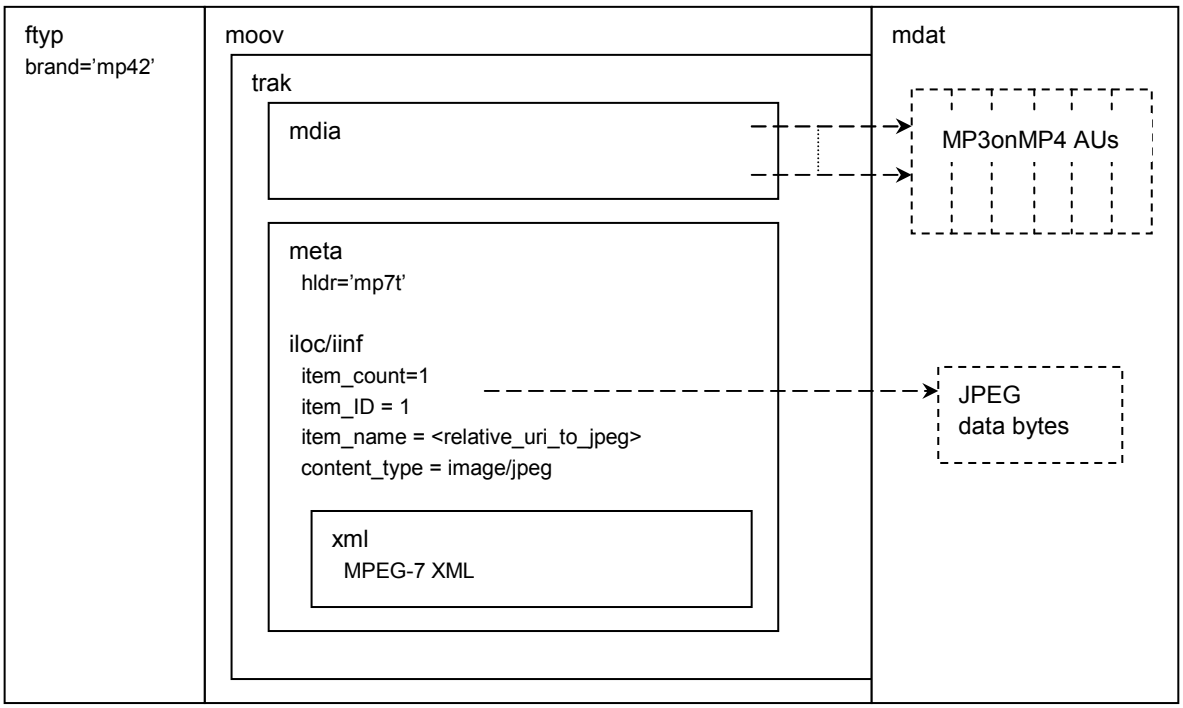

Figure 8 — Example of file structure for mp4 song file containing one audio track with MP3onMP4 audio, MPEG-7 meta-data and JPEG image

File Type: major-brand = 'mp42', compatible-brands = 'mp42', 'iso2', 'MMU2'.

Meta-data Handler Type: MPEG-7 Text (mp7t) at track level

Resource Lookup and Playback:

— A mp4 MAF application uses the *mdia* box and subsequent child boxes of the sample description to find and decode the MP3 data (stored as MP3onMP4 Access Units [14496-3:2005]).

— To present the optional JPEG Image, the application uses a combination of *iloc* and *iinf* boxes as shown below:

```
for(all items in the iloc) {
  if(iinf->content_type == image/jpeg) {
    //locate image using
        iloc->extent_offset
        iloc->extent_length
  }
}
```

NOTE 1 This structure contains MPEG-7 XML meta data pertaining to the mandatory track inside the mdat.

NOTE 2 Here the *meta* box inside the *trak* box is used to link the MP3 and the JPEG together. The coupling with the use of this box will suffice for single track MAFs.

5 Album and Playlist Format for the Music Player

5.1 Introduction

This clause describes an extension to the core song file format. It allows to create complete album files as well as playlist files that references external song files. The MPEG-21 File Format [ISO/IEC 21000-9:2005] and the MPEG-21 Digital Item Declaration (DID) [ISO/IEC 21000-2] is used to enable this functionality.

5.2 Single Track Album

The explanation of the structure starts with the special case of an album that contains only one single song.

An MPEG-21 file contains an MPEG-21 DID as entry point and a single, hidden mp4 song file that is built according to the previous Clause 4. The DID and the mp21 meta box are used to identify and locate the resources. The relationship between the mp4 song file, its MPEG-7 meta data and the optional JPEG image is held in the DID to give applications quick access to the meta-data of the album files (artist, title, cover image,…) without parsing into the hidden mp4 song file(s).

File Type: major-brand = ‘mp21’, compatible-brands = ‘mp21’, ‘MMU2’.

Meta-data Handler Type: MPEG-21 (mp21) at file level

Resource Lookup and Playback:

— The structure uses a parent *meta* box at the file level to refer to the hidden mp4 file inside the *mdat* box and directly to items (MPEG-7 meta data and JPEG image) inside the hidden *mp4 file*.

— A MAF application shall decode this file using the DID as an entry point to the resource mapping technique as described in the following stub code:

```
for (all Resources/Statement elements in a DID) {
        //Get ref attribute
       if( iinf-> item_name = ref->val && iinf->content_type = ref->mimeType) {
               //use extent_offset and extent_length from iloc to find the //chunk of bytes
        }
}
```


— To enable this resource mapping, the DID Resource element must refer directly to the MPEG-7 and JPEG resource in the hidden *mp4 file*. This means the offset and length for these items in the *iloc* box is worked out directly from the file level *meta* box, where any complications of having to decode offsets and lengths inside the hidden mp4 file *(moov->trak->meta)* are avoided.

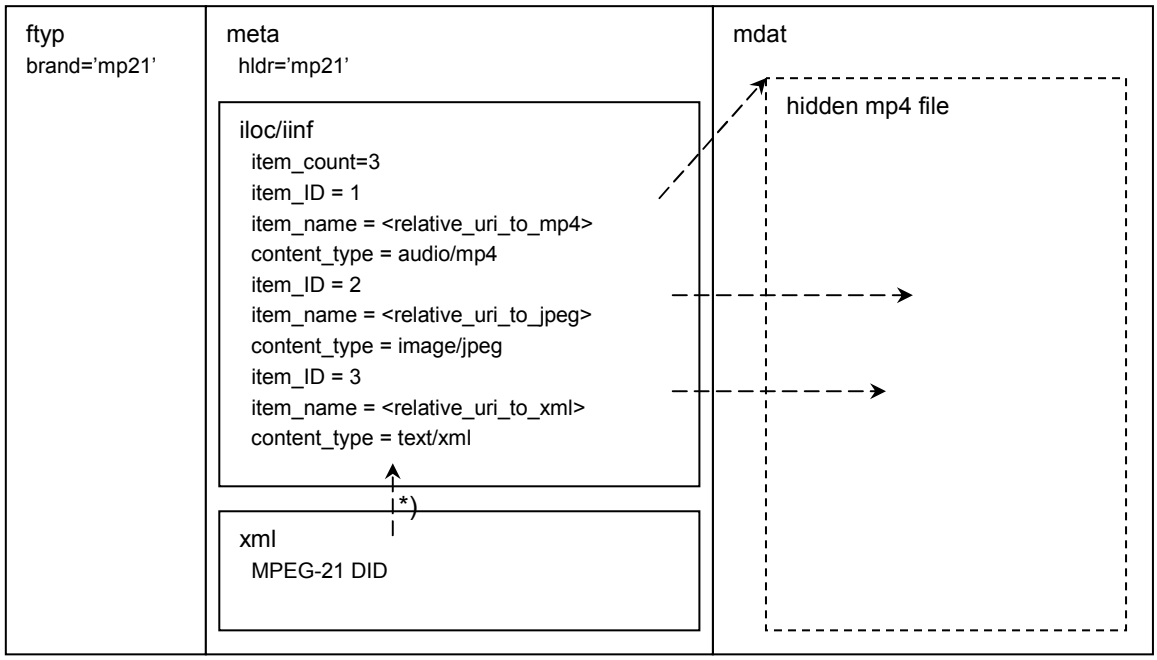

Figure 9 — Example of file structure for mp21 file containing one hidden mp4 song file

*) Arrows indicate that there is a resource mapping process from the DID to the hidden mp4 file