International Standard

ISO/IEC 21122-3

Third edition

Information technology — JPEG XS low-latency lightweight image coding system —

Part 3: Transport and container formats

PROOF/ÉPREUVE

Reference number
ISO/IEC 21122-3:2024(en)

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

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 third edition cancels and replaces the second edition (ISO/IEC 21122-3:2022), which has been technically revised.

The main changes are as follows:

— support for JPEG XS codestreams using temporal differential coding (TDC).

— clarifications on coding of interlaced signals.

A list of all parts in the ISO/IEC 21122 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

This document is part of a series of standards for a low-latency lightweight image coding system, denoted JPEG XS.

In many use cases during production or transmission of a movie, limiting the latency and the recompression loss is a more important aspect than the compression efficiency. The JPEG XS coding system offers compression and recompression of image sequences with very moderate computational resources while remaining robust under multiple compression and decompression cycles and mixing of content sources, e.g. embedding of subtitles, overlays or logos. Typical target compression ratios ensuring visually lossless quality are in the range of 2:1 to 20:1, depending on the nature of the source material. The end-to-end latency can be confined to a fraction of a frame, typically between a small number of lines down to below a single line.

This document specifies transport and container formats for JPEG XS codestreams. It also defines metadata that enriches transport protocols for transmission of image sequences, in order to facilitate transport, editing and presentation.

Information technology — JPEG XS low-latency lightweight image coding system —

Part 3:
Transport and container formats

1 Scope

This document defines transport and container formats for JPEG XS codestreams as specified in ISO/IEC 21122-1. It defines file formats for working with still image and motion image sequence files on computer platforms and gives guidance on how to embed the codestream in transport streams, allowing internet-based communication.

This document uses already existing specifications for file formats and extends them for the embedding of JPEG XS codestreams.

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

ISO/IEC 646, *Information technology — ISO 7-bit coded character set for information interchange*

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

ISO/IEC 11578, *Information technology — Open Systems Interconnection — Remote Procedure Call (RPC)*

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

ISO/IEC 21122-1, *JPEG XS low-latency lightweight image coding system — Part 1: Core coding system*

ISO/IEC 21122-2, *JPEG XS low-latency lightweight image coding system — Part 2: Profiles and buffer models*

ISO/IEC 23008-12:2022, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image File Format*

ISO/CIE 11664-1, *Colorimetry — Part 1: CIE standard colorimetric observers*

Rec. ITU-T H.273 | ISO/IEC 23091-2, *Coding-independent code points — Part 2: Video*

ANSI/CTA 861-G:2016, *A DTV Profile for Uncompressed High Speed Digital Interfaces*

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

3 Terms and definitions

For the purposes of this document the terms and definitions given in ISO/IEC 14496-12, ISO/IEC 21122-1, ISO/IEC 21122-2, ISO/IEC 23008-12 and the following 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
aux
auxiliary component channel typically used as opacity channel or alpha mask

3.2
big-endian
byte ordering from the most significant to the least significant byte of multi-byte value representations

3.3
box
structured collection of data describing the image or the image decoding process

3.4
box content
data wrapped within the *box* (3.3) structure

3.5
box type
kind of information stored with the *box* (3.3)

3.6
byte
group of 8 bits

3.7
coding-independent code point
code point based on enumerated values for the definition of the colourspaces

Note 1 to entry: Code points defined in Rec. ITU-T H.273 | ISO/IEC 23091-2.

3.8
high efficiency image file format
image file format which can embed still images and *motion sequences* (3.11)

Note 1 to entry: Based on ISO/IEC 23008-12.

3.9
image collection
unordered set of images without an implied or signalled presentation order or presentation time stamps

3.10
JXS
still image file format with JPEG XS compressed images

3.11
motion sequence
movie
timed sequence (3.15) of images

3.12
sample
<coding> single element in the two-dimensional image array which comprises a component

Note 1 to entry: This definition is used in Annex A.

3.13
sample
<ISOBMFF> all the data associated with a single time

Note 1 to entry: This definition is used in Annexes B and C as data associated with one coded image in a sequence.

3.14
superbox
box (3.3) that carries other boxes as payload data

3.15
timed sequence
linearly ordered sequence of media entities such as images where each entity is presented at a well defined time stamp

4 Symbols and abbreviated terms

4.1 Symbols

N_c	number of components in an image as defined in ISO/IEC 21122-1
P_{lev}	Level, sublevel and frame-buffer level (if applicable) a particular codestream conforms to as defined in ISO/IEC 21122-2
P_{pih}	profile a particular codestream conforms to as defined in ISO/IEC 21122-2
`Picture()`	JPEG XS codestream as defined in ISO/IEC 21122-1
`Codestream_Header()`	codestream header preceding the image data in the codestream without any `tpc_marker()` as defined in subclause A.5.5
`Codestream_Body()`	coded image data in the codestream including any `tpc_marker()` if such a marker is present, without `Codestream_Header()` as defined in subclause A.5.5

4.2 Abbreviated terms

For the purposes of this document the abbreviated terms given in ISO/IEC 14496-12, ISO/IEC 21122-1, ISO/IEC 21122-2, ISO/IEC 23008-12 and the following apply.

CICP	coding-independent code points
CIE	Commision Internationale de l'Eclairage
HEIF	high efficiency image file format
ISOBMFF	iso base media file format
LSB	least significant bit
MSB	most significant bit
UTF-8	8-bit Unicode transformation format as defined in ISO/IEC 10646

4.3 Naming conventions for numerical values

Integer numbers are expressed as bit patterns, hexadecimal values, or decimal numbers. Bit patterns and hexadecimal values have both a numerical value and an associated particular length in bits.

Hexadecimal notation, indicated by prefixing the hexadecimal number by "0x", may be used instead of binary notation to denote a bit pattern having a length that is an integer multiple of 4. For example, 0x41 represents an eight-bit pattern having only its second most significant bit and its least significant bit equal to 1. Numerical values that are specified under a "**Code**" heading in tables that are referred to as "code tables" are bit pattern values (specified as a string of digits equal to 0 or 1 in which the left-most bit is considered the most-significant bit). Other numerical values not prefixed by "0x" are decimal values. When used in expressions, a hexadecimal value is interpreted as having a value equal to the value of the corresponding bit pattern evaluated as a binary representation of an unsigned integer (i.e., as the value of the number formed by prefixing the bit pattern with a sign bit equal to 0 and interpreting the result as a two's complement representation of an integer value). For example, the hexadecimal value 0xF is equivalent to the 4-bit pattern '1111' and is interpreted in expressions as being equal to the decimal number 15.

5 Conformance

This document shares common definitions for the structure of files (a sequence of objects, called boxes here, and atoms in other similar file formats), and a common definition of the general structure of an object (the size and type) as specified in Annex A.

File formats representing either images, or image sequences shall follow the specifications in Annexes B, C and D. All these specifications require that readers ignore objects that are unrecognizable to them.

This document takes precedence over those on which it is based, in any case where there are differences or conflicts; however, no such conflicts are known to exist.

For better readability and understanding, the syntax description for the different file formats is done in the same way as in the base formats.

6 Colour specification

JPEG XS (as defined in ISO/IEC 21122-1) describes only the encoded bitstream of an image. The integrated multiple component transformation is only responsible for a decorrelation of the different colour components allowing for the reduction of the entropy in the data. In order to properly display or interpret the image, it is essential that the colourspace of that image data is properly characterized. For this purpose, the respective container format or transport channel signals the correct colourspace. The defined formats in this document for JPEG XS signals the colour space as specified in Rec. ITU-T H.273 | ISO/IEC 23091-2.

7 Organization of the document

Annex A specifies boxes and superboxes that can be used to signal metadata for isolated JPEG XS codestreams or sequencs of JPEG XS codestreams. The boxes are identical or similar to the boxes defined in other ISO standards, e.g. JPEG 2000 (Rec. ITU-T T.800 | ISO/IEC 15444-1). The boxes defined in this annex are used throughout all other annexes of this document.

Annex B defines the JXS file format for still images based on JPEG XS codestreams and the boxes specified in Annex A.

Annex C specifies the integration of JPEG XS codestreams in the ISOBMFF (as defined in ISO/IEC 14496-12) for use of image sequences as movie in a file format.

Annex D specifies the integration of JPEG XS codestreams in the HEIF file format (as defined in ISO/IEC 23008-12) allowing the integration of both still images as well as movies in one format.

Annex E specifies the Media Type registration for JPEG XS codestreams solely any file format container, i.e. not contained within the file formats specified by Annex B, Annex C or Annex D.

Annex A
(normative)

Syntax elements for JPEG XS compressed content

A.1 General

This annex defines syntax elements identifying and representing meta data of JPEG XS compressed images, image collections or image sequences. It forms the basis for file formats that applications may choose to wrap one or multiple JPEG XS codestreams. Annex B describes a concrete file format that represents an individual JPEG XS image. This specification is based on the same syntax as the box-based file format for JPEG 2000 in ISO/IEC 15444-1:2019, Annex I or ISO/IEC 15444-2:2021, Annex M.

This annex:

— specifies a binary container for image, image collections, image sequences and metadata;

— specifies a mechanism by which metadata (including vendor-specific information) can be included in files or transport streams specified by this document;

— specifies a mechanism to indicate image properties, such as the tonescale or colourspace of the image;

— specifies a mechanism by which readers can recognize the existence of intellectual property rights information in the file.

A.2 Specification of syntax elements

A.2.1 General

The syntax elements defined in this annex provide foundations for storing application specific data (metadata) in association with JPEG XS codestreams, such as information which is required to display the images or stream of images. As many applications require a similar set of information to be associated with the compressed image data, it is useful to define the format of that set of data along with the definition of the compression technology and codestream syntax.

Conceptually, the syntax elements specified in this annex encapsulate JPEG XS codestreams along with other core pieces of information about such codestreams. A file created from the syntax elements defined in this annex is loosely called a JPEG XS enabled file. However, this annex does not define a file format itself, it only provides syntax elements upon which a file format can be defined. A concrete file format based on this annex is specified in Annex B.

The building-block of all JPEG XS enabled files is called a box. All information is encapsulated in boxes. This annex defines several types of boxes; the definition of each specific box type defines the kinds of information that can be found within a box of that type. Some boxes will be defined to contain other boxes.

A.2.2 Greyscale, colour, multi-component specification

One of the most important aspects of a file format is that it specifies the colourspace of the contained image data. In order to properly display or interpret the image data, it is essential that the colourspace of that image is properly characterized. The syntax elements defined in this subclause provide one method to specify the colourspace of the image based on coding-independent code points (CICP). The CICP enumerated method specifies the colourspace of an image by the use of three numeric values that identifies the colourspace. The set of supported colourspaces is specified in A.5.4.3. The allowed values are a subset of the code points defined in Rec. ITU-T H.273 | ISO/IEC 23091-2.

A.2.3 Inclusion of auxilliary channels

In many applications, components other than the colour channels are required (auxilliary channels). For example, many images used on web pages contain opacity information; the browser uses this information to blend the image into the background. Another example is the use of alpha channels in video production; video mixers use this alpha channel to mix multiple images. It is thus desirable to include both the colour and auxiliary channels within a single codestream.

Syntax elements defined in Annex A provide means to indicate the presence of auxiliary channels (such as opacity), to define the type of these channels, and to specify the ordering and source of these. When a reader opens a JPEG XS enabled file, it determines the ordering and type of each component. The application shall then match the component definition and ordering from the JPEG XS enabled file with the component ordering as defined by the colourspace specification. Once the file components have been mapped to the colour channels, the decompressed image can be processed through any needed colourspace transformations.

How applications respond to opacity or other auxiliary channels is outside the scope of this document.

A.2.4 Metadata

One important aspect of the syntax elements defined in Annex A is the ability to add metadata to a JPEG XS enabled file.

Some of the boxes provide a set of tools by which applications can add vendor-specific information to the JPEG XS enabled file format, like the Exif box or the XML box. These boxes are optional in conforming files and may be ignored by conforming readers.

A.2.5 Temporal differential coding of interlaced signals

Each field of an interlaced frame constitutes an independent image in the sense of ISO/IEC 21122-1. Coding parameters of the two fields shall then match, i.e. if temporal differential coding is enabled for the top field, it shall also be enabled for the bottom field and vice versa. Temporal differential coding then predicts between identical fields, i.e. the top field shall be predicted from the last top field in the past, and the bottom field shall be predicted from the bottom field in the past.

Conceptionally, an interlaced encoded JPEG XS file can therefore be understood as two otherwise independent JPEG XS streams interleaved within each other, where one stream consists of all top fields and the other consists of all bottom fields.

A.3 Concept of boxes

A.3.1 Key to graphical descriptions

Each box is described in terms of its function, usage and length. The function describes the information contained in the box. The usage describes the logical location and frequency of this box in the file. The length describes which parameters determine the length of the box.

These descriptions are followed by a figure that shows the order and relationship of the parameters in the box. Figure A.1 shows an example of this type of figure. A rectangle is used to indicate the parameters in the box. The width of the rectangle is proportional to the number of bytes in the parameter. A shaded rectangle (diagonal stripes) indicates that the parameter is of varying size. Two parameters with superscripts and a grey area between them indicate a run of several of these parameters. A sequence of two groups of multiple parameters with superscripts separated by a grey area indicates a run of that group of parameters (one set of each parameter in the group, followed by the next set of each parameter in the group). Optional parameters or boxes are shown with a dashed rectangle.

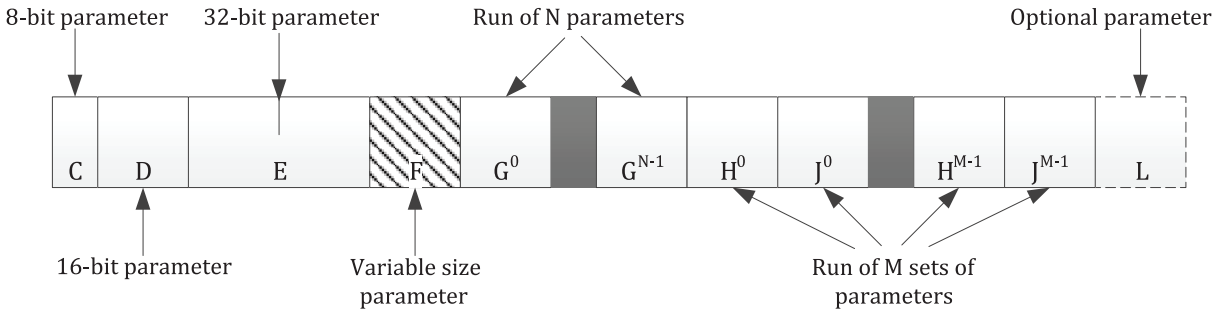

Figure A.1 — Example of a box description

The figure is followed by a list that describes the meaning of each parameter in the box. If parameters are repeated, the length and nature of the run of parameters is defined. As an example, in Figure A.1, parameters C, D, E and F are 8-, 16-, 32-bit and variable lengths, respectively. The notation G^0 and G^{N-1} implies that there are N different parameters, G^i, in a row. The group of parameters H^0 and H^{M-1}, and J^0 and J^{M-1} specify that the box will contain H^0, followed by J^0, followed by H^1 and J^1, continuing to H^{M-1} and J^{M-1} (M instances of each parameter in total). Also, the field L is optional and may not be found in this box.

After the list is a table that either describes the allowed parameter values or provides references to other tables that describe these values.

Some boxes may carry other boxes as payload data. Such boxes are denoted as superboxes. The payload size of a superbox is given by the sum of the box lengths of all the boxes it contains.

In addition, in a figure describing the contents of a superbox, an ellipsis (...) is used to indicate that the contents of the file between two boxes are not specifically defined. Any box (or sequence of boxes), unless otherwise specified by the definition of that box, may be found in place of the ellipsis.

For example, the superbox shown in Figure A.2 shall contain an AA box and a BB box, and the BB box shall follow the AA box. However, there may be other boxes found between boxes AA and BB. Dealing with unknown boxes is discussed in A.6.

Figure A.2 — Example of a superbox description

A.3.2 Box definition

Physically, each object in the file is encapsulated within a binary structure called a box. That binary structure is as in Figure A.3, and detailed in Table A.1.

Figure A.3 — Organization of a box

LBox Box length. This field specifies the length of the box, stored as a 32-bit big-endian unsigned integer. This value includes all of the fields of the box, including the length and type. If the value of this field is 1, then the XLBox field shall exist and the value of that field shall be the actual length of the box. If the value of this field is 0, then the length of the box was not known when the LBox field was written. In this case, this box contains all bytes up to the end of the file. If a box of length 0 is contained within another box (its superbox), then the length of that superbox shall also be 0. This means that this box is the last box in the file. The values 2-7 are reserved for ISO/IEC use.

TBox Box type. This field specifies the type of information found in the DBox field. The value of this field is encoded as a 32-bit big-endian unsigned integer. However, boxes are generally referred to by an ISO/IEC 10646 character string translation of the integer value. For all box types defined within this document, box types are indicated as both character string (normative) and as 4-byte hexadecimal integers (informative). Also, a space character is shown in the character string translation of the box type as "\040". All values of TBox not defined within this documentare are reserved for ISO/IEC use.

XLBox Box extended length. This field specifies the actual length of the box if the value of the LBox field is 1. This field is stored as an 64-bit big-endian unsigned integer. The value includes all of the fields of the box, including the LBox, TBox and XLBox fields.

DBox Box contents. This field contains the actual information contained within this box. The format of the box contents depends on the box type and is defined individually for each type.

Table A.1 — Binary structure of a box

Field name	Size (bits)	Value
LBox	32	0, 1, or 8 to (2^{32}– 1)
TBox	32	Variable
XLBox	64 0	16 to (2^{64}– 1); if LBox = 1 Not applicable; if LBox ≠ 1
DBox	Variable	Variable

For example, consider the illustration in Figure A.4 of a sequence of boxes, including one box that contains other boxes:

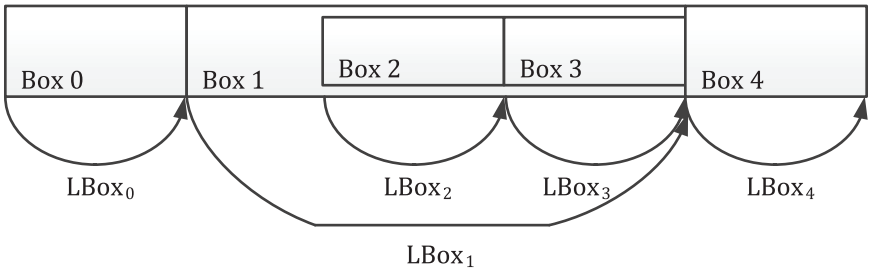

Figure A.4 — Illustration of box lengths

As shown in Figure A.4, the length of each box includes any boxes contained within that box. For example, the length of Box 1 includes the length of Boxes 2 and 3, in addition to the LBox and TBox fields for Box 1 itself. In this case, if the type of Box 1 was not understood by a reader, it would not recognize the existence of Boxes 2 and 3 because they would be completely skipped by jumping the length of Box 1 from the beginning of Box 1.

A.4 Overview of defined boxes

Table A.2 lists all boxes defined in Annex A. Indentation within the table indicates the hierarchical containment structure of the boxes within a JXS file as defined in Annex B.