INTERNATIONAL STANDARD

ISO/IEC 23000-6

First edition
2009-04-15

Corrected version
2010-09-01

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

Part 6:
Professional archival application format

Technologies de l'information — Format pour application multimédia (MPEG-A) —

Partie 6: Format pour application d'archivage professionnel

© ISO/IEC 2009

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 23000-6:2009
https://standards.iteh.ai/catalog/standards/sist/9475470e-87ec-43e2-9fc5-
f1ca9a52e864/iso-iec-23000-6-2009

COPYRIGHT PROTECTED DOCUMENT

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 23000-6 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

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 6: Professional archival 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*

— *Part 11: Stereoscopic video application format*

This corrected version of ISO/IEC 23000-6:2009 corrects the part title.

Introduction

The advance of digital multimedia technology has made the creation of digital multimedia content easier. This has resulted in an abundance of digital multimedia contents available for user consumption. It also necessitates ways to manage those digital multimedia contents. Digital multimedia contents should be well-handled and well-preserved in content archives so that part of or all of the contents aggregation can be reused for creation of new contents.

For preserving digital multimedia contents in an archive, one has to provide a packaging mechanism together with preservation technologies for data protection, data integrity, and data compression. In addition, the consumption of the archive can continue long after it has been created; therefore, the context information that describes the context of the archive and digital multimedia contents in it should also be contained in the package. This necessary context information may include information that can answer who, what, where, when, and why questions about the archive and the digital multimedia contents archived in it.

This part of ISO/IEC 23000 is an MPEG standard that specifies component technologies and their integration for digital multimedia content archive. ISO/IEC 23000 (also known as "MPEG-A"), is an MPEG standard defined by selecting readily tested and verified tools taken from the MPEG body of standards and combining them to form a AF (Multimedia Application Format). If a needed piece of technology is not provided within the MPEG, additional technologies originating from other organizations can be included by reference in order to facilitate the envisioned application format. For digital multimedia content archives, MPEG has designated this part of ISO/IEC 23000 "*Part 6: Professional archival application format.*" In other parts of this International Standard, the term "Professional Archival Application Format (PA-AF)" refers to this part of ISO/IEC 23000.

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

Part 6:
Professional archival application format

1   Scope

This part of ISO/IEC 23000 specifes the professional archival application format (PA-AF). The purpose of the PA-AF is to provide a standardized packaging format for digital files. This packaging format can also serve as an implementation of the information package specified by the reference model of the open archival information system (OAIS). The OAIS reference model is a framework for understanding and applying concepts necessary for long-term digital information preservation (where "long-term" is long enough to be concerned about changing technologies). In addition, PA-AF can also be used as an intermediate or exchange packaging format for any kind of multimedia content.

While a general archival process may include processes starting from creation, to delivery to archival system, and dissemination to consumers, PA-AF is limited in scope as follows. PA-AF does not specify how input content is created. PA-AF does not specify any agreement of how the content should be handled and delivered to the archiving process. PA-AF assumes that input content for the archiving process is available in an appropriate digital format. PA-AF specifies the format of a digital archive produced by the archival process. PA-AF does not specify how the archive output by the archival process is disseminated to end-users.

PA-AF specifies a metadata format to describe the original structure of digital files archived in a PA-AF file. PA-AF specifies a metadata format to describe context information related to a PA-AF file and digital files archived in it. PA-AF specifies a metadata format to describe necessary information to reverse the pre-processing processes applied to digital files prior to archiving them in a PA-AF file. PA-AF specifies a file format for carriage of the metadata formats and digital files.

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 14721:2003, *Space data and information transfer systems — Open archival information system — Reference mode*

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

ISO/IEC 14496-12:2008, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format (technically identical with ISO/IEC 15444-12)*

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

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

ISO/IEC 21000-3:2003, *Information technology — Multimedia framework (MPEG-21) — Part 3: Digital Item Identification*

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*

ISO/IEC 21000-14:2007, *Information technology — Multimedia framework (MPEG-21) — Part 14: Conformance Testing*

IETF RFC 1738, *Uniform Resource Locators (URL), IETF Request for Comments: 1738*, December 1994

IETF RFC 1808, *Relative Uniform Resource Locators, IETF Request for Comments: 1808*, June 1995

IETF RFC 3629, *UTF-8, a transformation format of ISO 10646, IETF Request for Comments: 3629,* November 2003

IETF RFC 3986, *Uniform Resource Identifier (URI): Generic Syntax, IETF Request for Comments: 3986*, January 2005

W3C XINCLUDE, *XML Inclusions (Xinclude) Version 1.0, W3C Recommendation*, 15 November 2006

W3C XML, *Extensible Markup Language (XML) 1.0 (Fourth Edition), W3C Recommendation, 16 August 2006*

W3C XML C14N, *Cannonical XML Version 1.0, W3C Recommendation*, 15 March 2001

W3C XML NAMES, *Namespaces in XML 1.0, W3C Recommendation*, 16 August 2006

W3C XML SCHEMA, *XML Schema 1.1 Part 1: Structures (Second Edition), W3C Recommendation,* 28 October 2004

W3C XML SCHEMA, *XML Schema 1.1 Part 2: Datatypes (Second Edition), W3C Recommendation,* 28 October 2004

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 14721:2003, ISO/IEC 14496-3, ISO/IEC 14496-12:2008, ISO/IEC 14496-14:2003, ISO/IEC 21000-2:2005, ISO/IEC 21000-3:2003, ISO/IEC 21000-4:2006, ISO/IEC 21000-5:2004, ISO/IEC 21000-9:2005, ISO/IEC 21000-14:2007 and the following apply.

3.1
application-specific context information
context information whose semantics are not defined in the general PA-AF specification but may be used in specific application domains

3.2
archive structure information
information that describes structural relationships among **content information**

3.3
content information
file or fragment of file that is archived in a **PA-AF file**

3.4
context information
information that describes the context of a **PA-AF file** and **content information** stored in it

3.5
file attribute
information that preserves attributes attached to an input file by the file system before it is archived in a **PA-AF file**

3.6
information package
package containing the file(s) being archived and associated descriptioin information needed to aid in the preservation of the archived file(s) in a specific file format

3.7
professional archival application format file
PA-AF file
archive whose format complies with the PA-AF specification

3.8
pre-processing tools
software or software module used for processing before a file is archived in a **PA-AF file**

NOTE   Pre-processing tools include tools for lossless data compression, encryption, governance checking, and integrity checking.

3.9
pre-processing information
information carried in a **PA-AF file** describing tool(s) and other settings required to reverse pre-processing action(s) applied to **content information**

3.10
preservation description information
information describing **content information** stored in a **PA-AF file**

NOTE   It consists of archive structure information, context information, and pre-processing information.

4   PA-AF Overview

PA-AF archives digital files in a PA-AF file. In addition to containing digital files being archived, a PA-AF file also contains information for the preservation of the archived digital files. While implementation of a packaging tool that complies with PA-AF is out of the scope of this specification, Figure 1 outlines an informative packaging tool that may produce an output file that complies with a PA-AF file. The tool consists of the following modules:

— file structure information generator, which analyzes and generate meta information to model hierarchical structure of the input files;
— context information generator, which creates meta information to record context information related to the output PA-AF file and input files to be archived;
— pre-processing information generator, which creates meta information for required tools and their execution parameters to reverse any pre-processing processes applied to the input files;
— archive header wrapper, which combines all the generated meta information into PA-AF file header;
— file formatter, which takes the header and input files (original or after pre-processed) and wraps them in a file.
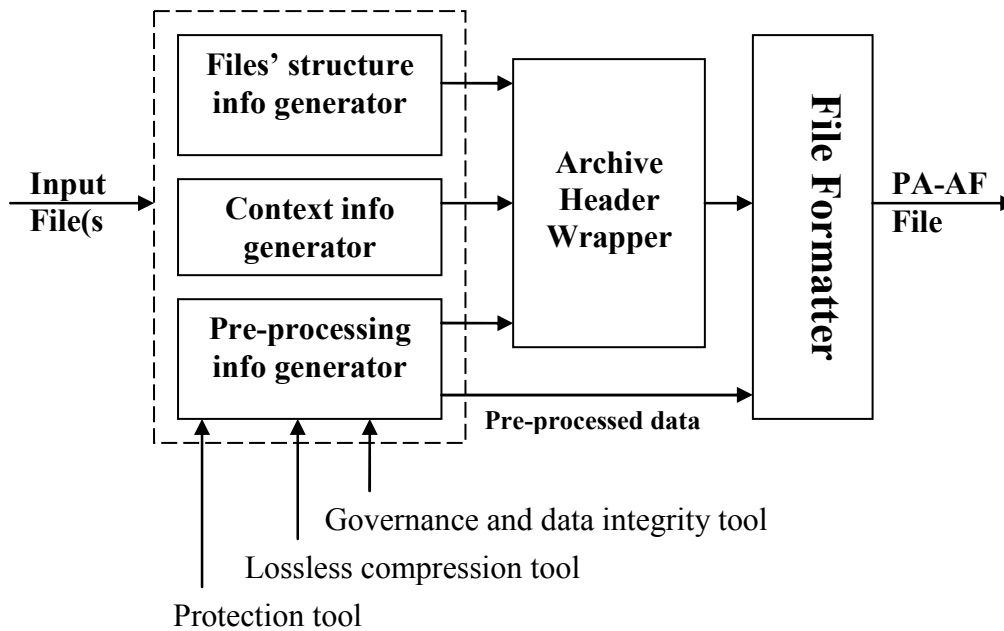
Figure 1 — Overview of a PA-AF packaging tool for creating a PA-AF file

5   Normative Components of PA-AF File

A PA-AF file consists of header and content part as illustrated in Figure 2. The header part contains information, called Preservation Description Information, needed to understand the PA-AF file itself and all files archived in it. The content part contains one or more archived files. These archived files are called Content Information.
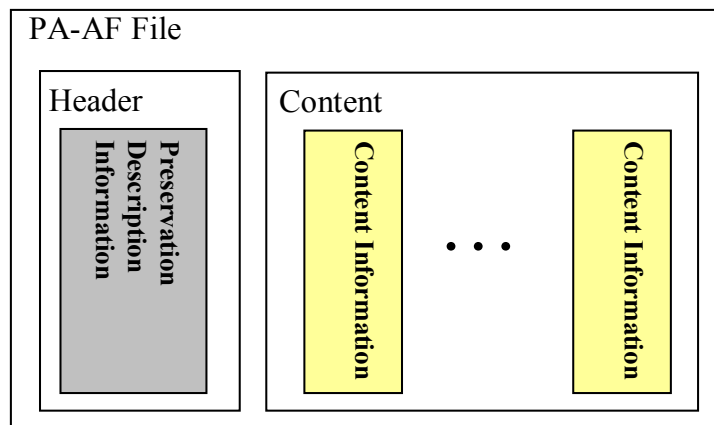
Figure 2 — Logical view of a PA-AF file

Content Information includes digital data in its original format as input into the PA-AF file and/or in the format after pre-processed with pre-processing tools allowed by this specification. The categories of pre-processing tool are lossless data compression, reversible data protection such as encryptions, and removable data attached to the content for usage governance and data integrity, such as checksums, and digital licenses.

Preservation Description Information is in XML metadata format and shall comply with W3C recommendations related to XML. It includes Archive Structure Information, Context Information, and Pre-processing Information.

Archive Structure Information describes structural relationships among Content Information. Figure 3 illustrates the importance of Archive Structure Information. While the original structure of digital files input into a PA-AF file can be hierarchical in one or more directories, in a physical PA-AF file format, they are stored in flat manner. Archive Structure Information preserves the original hierarchical structure of the input digital files. When Content Information is extracted from the PA-AF file, the structure of output digital files is as it was input into the PA-AF. Archive Structure Information also acts as an entry point to access Context Information and Pre-processing Information in the Content Information.

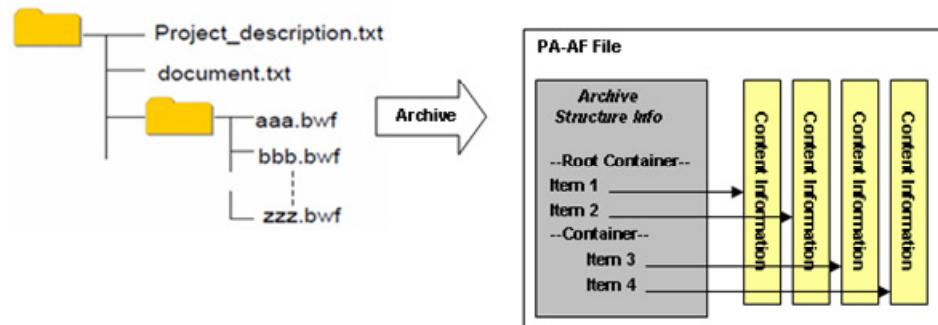**Figure 3 — Usage of Archive Structure Information**

Context Information describes context information attached to a PA-AF file and Content Information.

It includes:

— creation information of the PA-AF file and content Information, such as information about what, how, when, where, who, and why;

— profile information of the Content Information, such as the file format, file size, audio and visual profile of the content (if it is an audio visual data);

— access history to the Content Information, which records any actions applied to the Content Information, such as archiving and extraction; this record may include who the actor is and when the action is performed;

— application specific context information.

An application domain that adopts PA-AF specification could add additional information to describe Content Information in the PA-AF file to satisfy its specific requirements by means of application specific context information. Figure 4 illustrates how PA-AF accommodates such a mechanism. Application- specific context information can be manifested in any meta information format and archived in a PA-AF file as Content Information. PA-AF provides a link to refer to this application-specific context information so that an application that can understand this information can read and use it.
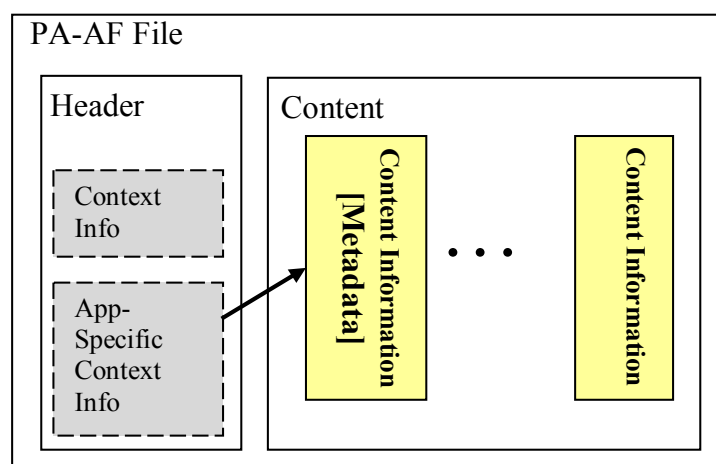

Figure 4 — Application-specific context information in a PA-AF file

Pre-processing Information describes profile of tools that can be used to reverse pre-processing applied to Content Information. It contains information, such as identification of the tool, required parameters to execute the tool, and part of Content Information pre-processed with that tool, and where to acquire the tool.

Required components technologies to meet normative components of PA-AF are as follows:

— MPEG-21 DIDL 2nd Edition Profile for PA-AF

— ISO/IEC 21000-3:2003

— MPEG-21 Intellectual Property Management and Protection Components Base Profile for PA-AF

— ISO/IEC 21000-5:2004/Amd.1:2007

— MPEG-7 Multimedia Descrption Scheme Profile for PA-AF

— MPEG-21 File Format for PA-AF

— Lossless compression tool identifiers

— Encryption, hash, and digital signature identifiers

— Additional metadata dedicated for use in PA-AF only

The component technologies listed above can be used in combination to fulfil basic functionality and enhanced functionality of PA-AF as illustrated by Figure 5. The enhanced functionality is optional and implemented on top of the basic functionality. For example, the combination of ISO/IEC 21000-9:2005, MPEG-21 DIDL 2nd Edition Profile for PA-AF, ISO/IEC 21000-3:2003, and MPEG-7 Creation Information Tool of ISO/IEC 15938-5:2003 provides solutions to satisfy the basic functionality of PA-AF which is packaging Content Information in a PA-AF file. By adding MPEG-21 Intellectual Property Management and Protection Components Base Profile for PA-AF, one can add functionality, such as compression, protection, and integrity checking to the PA-AF. By adding ISO/IEC 21000-5:2004/Amd.1:2007, one can add license information to govern the usage of the PA-AF file. Finally, by adding MPEG-7 Multimedia Descrption Scheme Profile for PA-AF, one can have interoperable description of Content Information that can be exploited to implement functionality for interoperable content searching. The combination of all component technologies provides a full solution for PA-AF.

Data protection, compression, and integrity	Usage governance	Interoperable content searching	Application specific context info.
Basic Function **[Package input digital files into PA-AF archive]**			
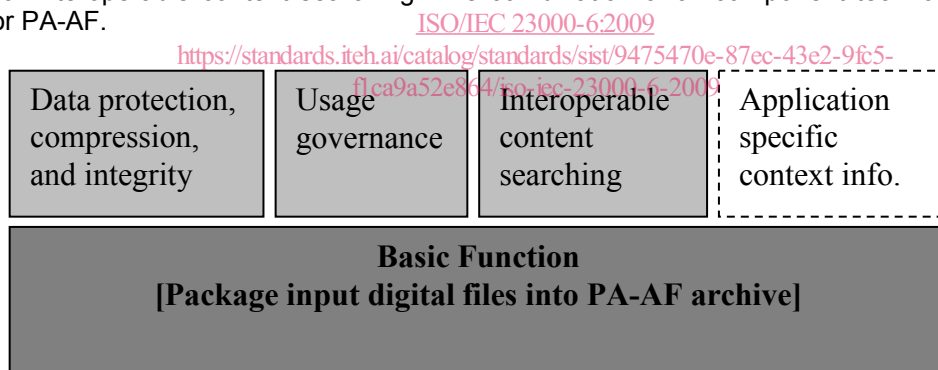
Figure 5 — Basic and enhanced functionality of PA-AF

6 Archive Structure Information

6.1 Overview

Archive Structure Information consists of a file structure model and file attribute model. The file structure model provides a description to preserve the original structure of input files to be archived, whereas the file attribute model provide a description to preserve the original description given by the file system to the file, such as file access information, system timestamp, etc. 6.2 provides explains the file structure model and explains the file attribute model.

6.2   MPEG-21 DIDL 2nd Edition Profile for PA-AF

MPEG-21 DIDL 2nd Edition Profile for PA-AF is a subset of ISO/IEC 21000-2:2005, with restrictions, and two additional attributes. The profile is designed to provide the following functionalities:

— modeling of original structure of input files for archive creation;

— provision of a pointer to the location of archived file in archive container;

— provision of a container for other information related to the archive itself and the Content Information.

The elements of MPEG-21 DIDL 2nd Edition Profile for PA-AF, their use, and restrictions applied on them are listed in Table 1.

Table 1 — Elements of MPEG-21 DIDL 2nd Edition Profile for PA-AF

MPEG-21 DIDL 2nd Edition Profile Element	Restriction
DIDL	Possible child element is only Container, and the cardinality is exactly 1
Container	As in 7.2.8 of ISO/IEC 21000-2:2005
Item	Possible child elements are Component and Descriptor
Component	Possible child element is Resource
Resource	Restricted to have no child element
	Possible attributes are mimeType, ref, and contentEncoding
	Resource element shall not carry inline binary data as its element content
Descriptor	Possible child element is only Statement
Statement	Possible attribute is only mimeType with value "text/xml"

The association of elements of MPEG-21 DIDL 2nd Edition profile for PA-AF and file structure is listed in Table 2.

Table 2 — Elements of MPEG-21 DIDL 2nd Edition Profile for PA-AF and file structure association

Element Name	Association to file structure
Container	Directory
	One directory is modeled by one Container element
Item	File
	One file is modeled by one Item element
Component	Fragment of file
	If a file is not fragmented, only one Component shall occur as child element of Item
	If a file is fragmented, the following are applied: – One Component is associated with one fragment of file – Occurance order of Component elements reflects order of fragments of file
Resource	Provide link to physical location of archived files in the archive container

The DIDL element shall contain exactly one Root Container. The Root Container can contain one or more Container element(s), which is/are associated with the directory, and/or one or more Item element(s), which is/are associated with the file.

If file is split into several fragments, each fragment shall be archived in separate PA-AF files. PA-AF does not mandate how to split files; however, it is recommended that the file be split according to its byte order. For example, if file is split into three fragments, then fragment 1 shall contain the data subset from byte 0 until byte X-1, fragment 2 shall contain the data subset from byte X until byte Y-1, and fragment 3 shall contain byte Y until the last byte of the original file. For every PA-AF file, the Item element associated to fragmented file shall have as many child Component elements as the number of file fragments. Annex G provides a showcase of fragmented files in PA-AF and how they are described.

6.3 File Attribute Model

The file attribute model is metadata designed to carry any attributes that can be attached to a directory or file in the file system. In use with MPEG-21 DIDL 2nd Edition profile for PA-AF, file attribute model metadata shall be carried under the Statement of first Descriptor of each Container or Item. US-ascii or UTF-8 as defined in IETF RFC 3629 shall be used for the character set of the default encoded path.

Figure 6 shows the structure of file attribute model metadata. Table 3 describes the semantics of file attribute model metadata. In addition to normative file attributes, PA-AF also provides placeholders to carry non-normative file attributes. These non-normative file attributes can be defined by the creator of a PA-AF file for use in specific application domains only.
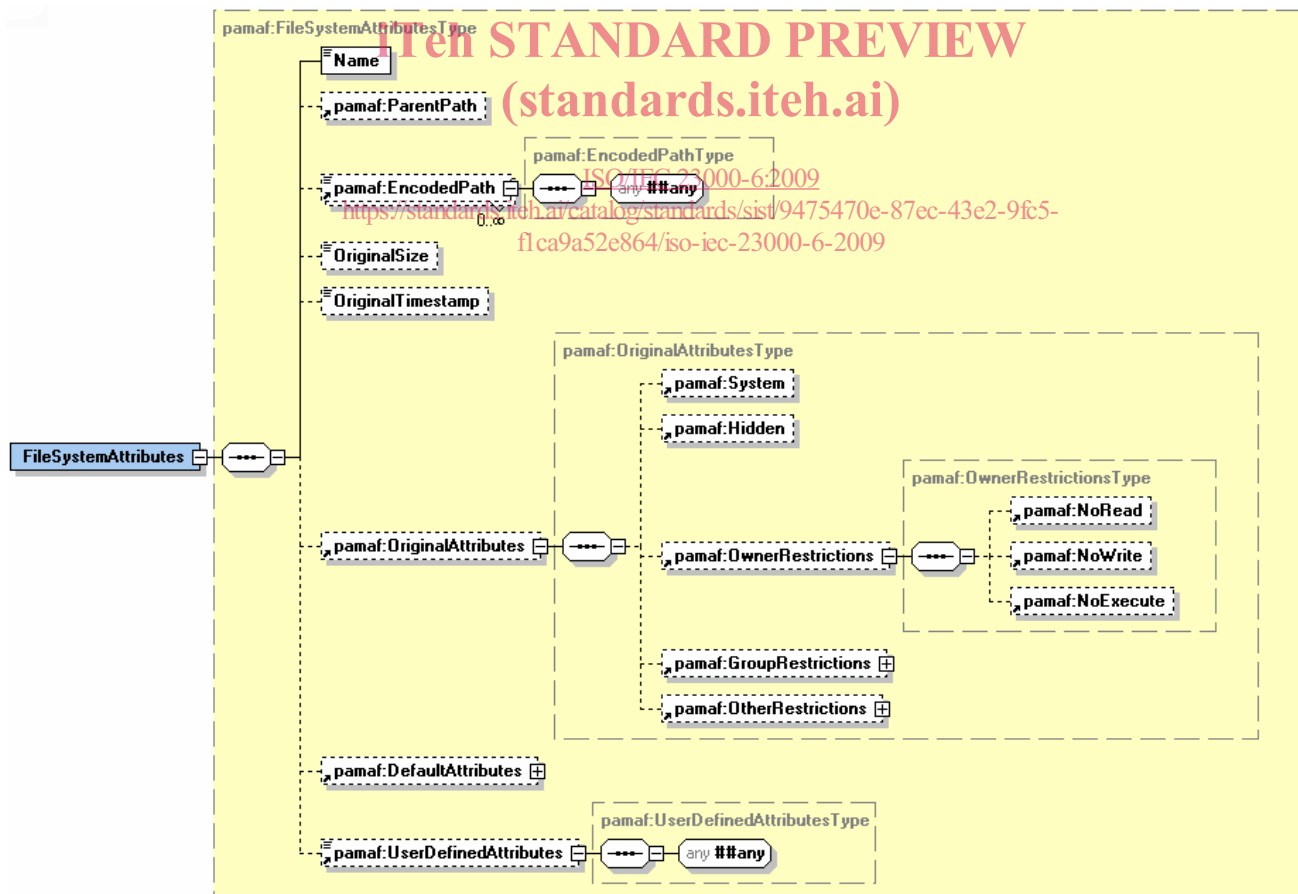

Figure 6 — Structure of file attribute model metadata

Table 3 — Semantics of file attribute model metadata

Name	Definition
FileSystemAttributes	Describes attributes that can be attached to a file or a directory of a file system
Name	Describes the name of the file or directory it describes
ParentPath	Describes the URI of absolute path from the file system root to the file or directory Only files and directories under the first Container needs to declare ParentPath; the remaining parent path of file or directories can be inferred.
EncodedPath	Describes the original path of the file encoded by base64 binary.
@charset	Attribute of EncodedPath. Describes the original character set that is used to encode the encoded path of the file. PA-AF adopts the character sets recommended by IANA Matrix which can be found at http://www.iana.org/assignments/character-sets and http://www.iana.org/assignments/media-types-parameters.
@original	Attribute of EncodedPath. Describes whether the encoded path is the original encoded path for the file. A file can have its path encoded in many encoded paths but may have only one original encoded path. The PA-AF extractor shall first try to restore the file in the path according the original encoded path in order to restore it to its original form. If this is not possible, it can try to use other available encoded paths.
@default	Attribute of EncodedPath. Describes whether the encoded path is the default encoded path for the file. A file can have its path encoded in many encoded paths but may have only one default encoded path. The PA-AF extractor shall first try to restore the file in the path according the original encoded path. if this is not possible, it can try to use other available encoded paths. If there is no encoded path that can be used for the target platform, the PA-AF extractor should generate a path name based on this default encoded path. The character set of the default encoded path should be US-ASCII or UTF-8 in order to allow PA-AF extractors to have at least one possible encoded path name.
OriginalSize	Describes the size of file. The unit is "byte"
OriginalTimestamp	Describes the original timestamp of a last modification time of the file or folder given by the file system. OriginalTimestamp is stored in the format "timePointType" adopted from MPEG-7 data types for universal representation across different operating systems and filing systems
OriginalAttributes	Describes attributes whose values are already set from its original filing system. The attributes includes "System", "Hidden", "OwnerRestrictions", "GroupRestrictions", and "OtherRestrictions"