INTERNATIONAL STANDARD

ISO 22311

First edition
2012-11-15

Societal security — Video-surveillance — Export interoperability

Sécurité sociétale — Videosurveillance — Interopérabilité de l'export

Reference number
ISO 22311:2012(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) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member 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 shall not be held responsible for identifying any or all such patent rights.

ISO 22311 was prepared by Technical Committee ISO/TC 223, *Societal security*.

Introduction

Video-surveillance is a crucial asset in intelligence collection, crime prevention, crisis management, forensic applications etc. The minimum requirement in societal security is for the authorities to be able to rapidly use the data collected by different CCTV systems from given locations.

This International Standard provides an export interoperability profile which constitutes the exchange format and minimum technical requirements that ensure that the digital video-surveillance contents exported are compatible with the replay systems, establish an appropriate level of quality and contain all the context information (metadata) necessary for their processing.

It is crucial for societal security that present and future video-surveillance systems implement this interface to allow efficient forensic processing of the material produced, often in massive quantities.

This International Standard also contains provisions to ensure that privacy measures can be implemented to protect the rights of the individuals.

This International Standard does not impose implementation methods or technological solutions. It relies heavily on individual technical standards separately developed and concentrates on minimum necessary profiles or subsets thereof to achieve its societal security objectives.

This International Standard is a blend of profiles of standards and practices, which combined, will achieve a minimum level of interoperability.

This implementation has only been possible because of standards produced by the following bodies:

— ISO/IEC JTC 1/SC 29/WG 11, Coding of moving pictures and audio (MPEG);

— ISO/IEC JTC 1/SC 29/WG 1, Coding of still pictures (JPEG);

— IEC/TC 79, Alarm systems and electronic security (including its European equivalent CENELEC/TC 79, Alarm systems and electronic security);

— ITU, International Telecommunication Union;

— IETF, Internet Engineering Task Force;

— SMPTE, Society of Motion Picture and Television Engineers;

— NATO, Standardization Agency.

The normative Annex A contains a metadata dictionary.

The importance of having images stored and presented to the user in such a way that their use is facilitated is presented in the informative Annex B.

Societal security — Video-surveillance — Export interoperability

1 Scope

This International Standard is mainly for societal security purposes and specifies a common output file format that can be extracted from the video-surveillance contents collection systems (stand alone machines or large scale systems) by an exchangeable data storage media or through a network to allow end-users to access digital video-surveillance contents and perform their necessary processing. The means of exchange are not part of this International Standard.

This common output file format relies on a combination of several technical standards that individually are not restrictive enough to provide the requested interoperability. These standards are formally referenced to avoid duplications or divergence. When appropriate to improve the interoperability, subsets or a limited number only of these standards are called.

Since video-surveillance recording often includes taking records of citizens, requirements relating to privacy, use of the records and their disposal are also considered.

Based on the above mentioned technical standards, the following format components are covered:

— Video;

— Audio;

— Metadata:

  — Descriptive (location, camera identifier, etc.)

  — Dynamic (date, time, pan, tilt, zoom, identification results, etc.)

— Encapsulation/packaging for the output file;

— Data/access security and integrity;

— Provisions for privacy;

— Informative data regarding the presentation to users.

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 10918-1:1994, *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines — Part 1*

ISO/IEC 10918-5, *Information technology — Digital compression and coding of continuous-tone still images: JPEG File Interchange Format (JFIF) — Part 5*[1)]

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

ISO/IEC 14496-3:2009, *Information technology — Coding of audio-visual objects — Part 3: Audio*

1) To be published.

ISO/IEC 14496-10:2012, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-12:2012, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

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

ISO/IEC 14496-15:2010, *Information technology — Coding of audio-visual objects — Part 15: Advanced Video Coding (AVC) file format*

ISO/IEC 15444-1:2004, *Information technology — JPEG 2000 image coding system: Core coding system — Part 1*

ISO/IEC 23000-10, *Information technology — Multimedia application format (MPEG-A) — Part 10: Surveillance application format*[2)]

IEC 62676-1-1, *Video surveillance systems for use in security applications — Part 1-1: Video system requirements*[3)]

IEC 62676-2-3, *Video surveillance systems for use in security applications — Part 2-3: Video transmission protocols — IP interoperability implementation based on web services*[4)]

ITU-T/Rec G.711, *Pulse code modulation (PCM) of voice frequencies*

SMPTE RP210.11-2008, *Metadata Dictionary Contents*

SMPTE 335M-2001, *Metadata Dictionary Structure*

SMPTE 336M-2007, *Data Encoding Protocol Using Key-Length Value*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
video-surveillance
surveillance by video means

3.2
forensic
related to or used in courts of law

NOTE This applies to video-surveillance used to produce legal evidence.

3.3
metadata
information to describe audiovisual content and data essence in a format defined by ISO or any other authority

EXAMPLE Time and date, text strings, location identifying data, audio and any other associated, linked or processed information.

3.4
static metadata
data associated with a digital image aside from the pixel values that does not change over time (or at least does not change over the addressed sequence)

2) To be published.

3) To be published.

4) To be published.

3.5
dynamic metadata
data associated with a digital image aside from the pixel values, which can change for each frame of a video sequence

3.6
CCTV system
surveillance system comprised of cameras, recorders, interconnections and displays that are used to monitor activities in a store, a company or more generally a specific infrastructure and/or a public place

3.7
logical structure
arrangement of data to optimize their access or processing by given user (human or machine)

3.8
geo-location
specific location defined by one of several means to represent latitude, longitude, elevation above sea level, and coordinate system

NOTE Geo-location generally means the meaningful specification of the position of a point or object on the earth. The term itself does not carry a prescription of the coordinate system to be used. Additional attributes associated with a geo-location are not a part of a geo-location specification.

3.9
scene location
collection of geo-locations that defines the perimeter of the viewable scene of a camera

NOTE The coordinate system is the same for each geo-location in the collection. There is at least one geo-location in the scene location. The geo-locations are ordered in either clockwise or counter clockwise order. Single geo-location scenes interpret the geo-location as the centre of the scene.

4 General

4.1 Concept

4.1.1 Video-surveillance systems generic architecture

A CCTV system usually consists of hardware, software and human elements.

A CCTV system for security applications presented as functional blocks, which portray the various parts and functions of the system, as well as the interactions with the human stakeholders is outlined in Figure 1.

This International Standard specifies the packaging and format of the data exchanged between the video and system management functional blocks of a CCTV system and societal security end-users external to the system as covered by the dotted zone of Figure 1.

Informative data regarding the presentation to users are given in Annex B.

The interactions between the components of the different functional modules of a CCTV system are specified by the IEC 62676 series of normative documents. This International Standard is an implementation of the provisions of 6.1.3 in IEC 62676-1-1 (to be published), which calls for publication of all necessary information allowing intended usage of the data produced. Accordingly this International Standard introduces the necessary extra requirements to the IEC 62676 series, without being in contradiction with them.

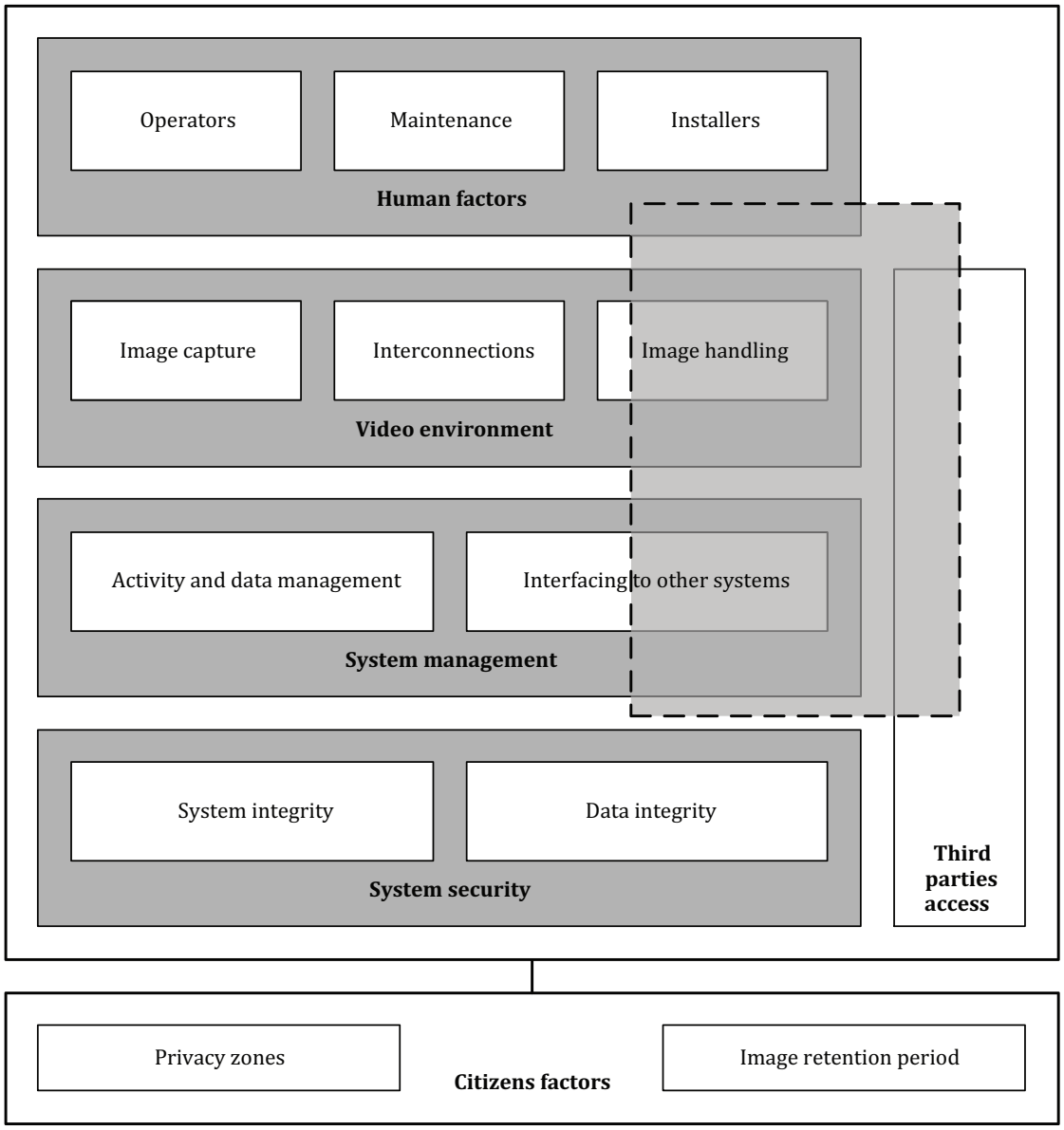

Figure 1 — Functional blocks of a CCTV system for security applications

4.1.2 Minimum requirements for interoperability

Societal security supposes interoperability of digital video-surveillance systems; this International Standard defines the minimum requirements applicable to the file formats used to export the collected contents (video, audio and associated metadata) to achieve this interoperability[5)].

All collected information shall be referenced to Coordinated Universal Time (UTC).

This format shall allow the file export of time slices of data coming from a selection of sources and preserve the time correlation between the contents, whatever export process (removable media or data transmission) is used.

The format shall be such that compatible, comparable processing of files exported by different systems (covering a same scene) with a common time base will be possible.

5) Pending legal authorization, this International Standard may allow streaming contents formatted as above, from a compliant CCTV system towards an external law enforcement facility, as its request or pushed by the CCTV system to allow hand-over in crisis situations.

Provisions will allow for the implementation of the applicable security, integrity and privacy protection measures.

As detailed hereafter, this export file format relies on several requirements for the elementary lower level formats and related protocols, respectively:

— Video;

— Audio;

— Metadata:

  — Static (localization, camera identifier, etc.)

  — Dynamic (date, time, pan, tilt, zoom, identification results, etc.)

— Container structure to integrate the above;

— Data security and integrity;

— Provisions for privacy.

The implementation of this International Standard shall be such that widely available Operating System (OS) independent tools will allow for minimal processing of received standard files by societal security organizations, such as forensic investigators, ensuring as a minimum the following and any combination thereof:

— Videos and metadata display;

— Direct access to the metadata without display of the videos;

— Selection of content time slots;

— Access to the sources defined by name or scene-location.

4.2 Relation with other standards

The different formats and prescriptions mentioned above are to a large extent subsets of existing standards; as a result, this International Standard is a blend of profiles of standards and practices which combined will achieve the expected minimum level of interoperability.

5 Requirements

5.1 General

The minimum requirement for a video-surveillance system to comply with this International Standard is to produce export files based on the following requirements.

These export files are essentially self-contained Audio-Video Packages which contain all information necessary to use the data, including location and time of all the scenes, without requiring access to the source system.

NOTE The requirements detailed hereafter are compatible with the implementation of the IEC 62676 series that make use of the data and of the stream payloads defined therein.

5.2 Requirements for the Audio-Video Package

5.2.1 General

The "Audio-Video Package" shall consist of a structure container, made of an indexing descriptor document (XML) and a collection of video and audio data, as well as metadata files. All of these files shall be in one hierarchical folder (or file system as may be the appropriate terminology for the operating system under which they are created).

A folder shall contain any number (and without any limitation) of the individual files described below and will cover an unlimited number of elementary contiguous time slots.

For easy random access to information, data shall be split into Directory Time Slots (DTS) and File Time Slots (FTS), where time is given in Coordinated Universal Time (UTC), as for GPS. Duration of a DTS is not critical; for operator convenience (typically, fast search of events by investigators), this duration shall be constant for a system and a new DTS shall be initiated at least once per day, starting with one hour.

As a minimum, any exported package shall contain once the descriptor of each of the Directory Time Slots covering (partially or in full) the corresponding time slot. This descriptor shall contain a time-stamped table of the changes that may have occurred in the configuration since the beginning of the DTS as well as the list of the contained FTS with their start time.

The duration of a File Time Slot (FTS) shall be comprised between 1 and 600 s. Within a time slot indexes shall allow to point accurately to any specific frame and time.

NOTE 1 It is allowed to zip or tar the package (again according to the operating environment of the source) and to transport it by any convenient means including placement on a portable storage media or file transfer protocol. These processes are not a part of this International Standard. Only the format of the package is covered in this International Standard.

An Audio-Video Package shall accordingly be organized in a structure as figured below with, in one top folder, a collection of time slices, arranged in DTS and FTS, of the relevant audio, video and other dynamic data contents and for each of them a description giving all available information relative to the corresponding time slice. Such a generic package organization is detailed in Figure 2.

For practical reasons, data naturally belong to groups, typically corresponding to a physical infrastructure, to data collected by a same Network Video Recorder (NVR) or to an administrative district; relationship with such groups shall be part of the description.

As a minimum interoperability condition, the naming rule shall be: DTS_info.xml.

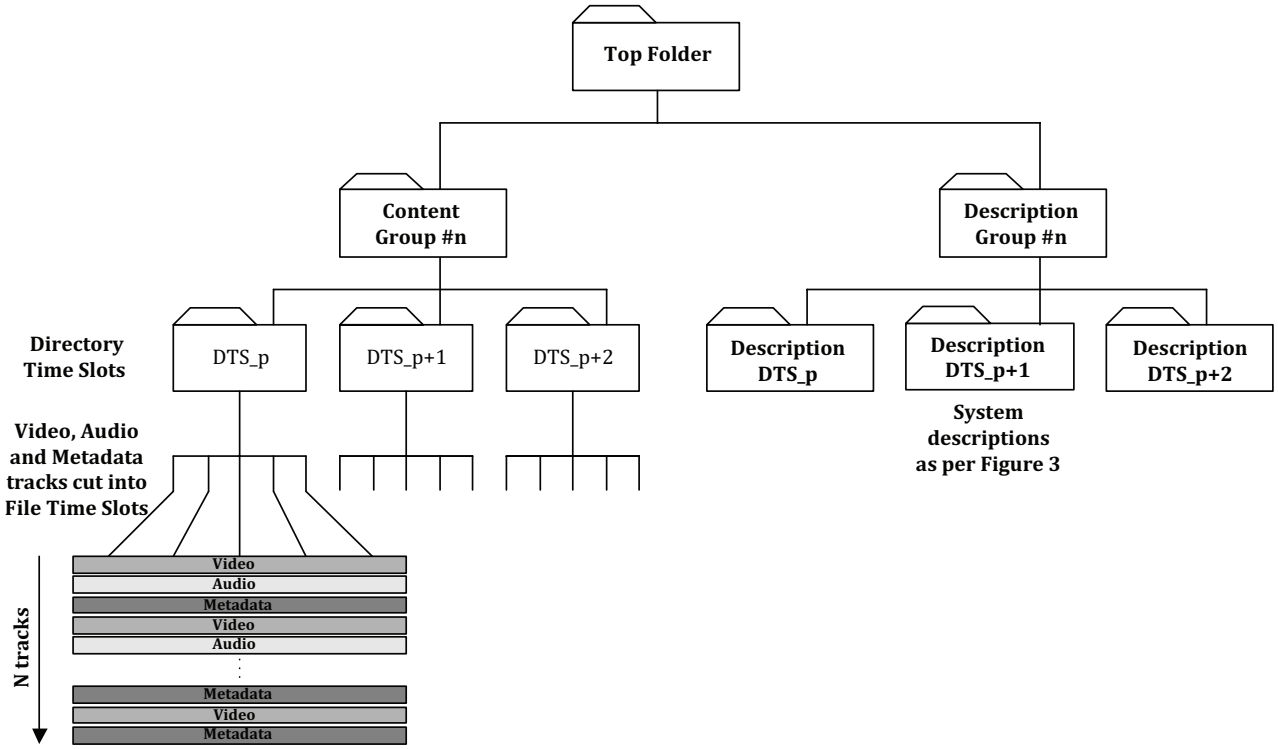

Figure 2 — Generic files organization

The Audio-Video Package XML Descriptor shall have the structure defined in Figure 3 below, which graphically describes a package with a group of contents at a given time.

This XML description shall be of "XSDcomplexType", naturally self-explanatory; it will describe the Group and all the sources belonging to the Group; the descriptions typically rely on the data listed in each subset. All the data do not need to be provided, but enough information shall be given to define without ambiguity all the Audio/Video contents contained in the folder. In a similar manner more information or content than explicitly mandated by this International Standard can be incorporated, provided it is properly described.

As detailed in 5.2.4, for a better interoperability, some information is mandatory and restrictions may apply to some of the listed items.

The «XSDelements» below shall be extracted from the XML Metadata scheme and namespace defined in IEC 62676-2-3. As a minimum interoperability condition, naming rule shall be:

— Directories: DIR_ < time stamp > (example: DIR_4E37B984)

— Files:<trackname>_<datatype>_<typeofencoding>_<timestamp>.<extension>(example:TESTVI20_video_H264_4E411AEC.mp4 orTESTVI50_video_H264_4E411AEC.d)

where

— time stamp is the hexadecimal DTS aligned representation of the file time in seconds as of 1 January 1970 (see Annex A Dates and Times section);

— track name is a chain of characters (0 to 9 and a to Z);

— data type is video, audio or metadata;

— type of encoding is, e.g. H264 or MPEG-2 for video, KLV or XML for metadata;

— the "d" extension shall be arbitrarily used where the "mp4" extension cannot be used.

The XML description of the group shall be of "XSDcomplexType" and typically consist in its location and its contact points.

The descriptive information shall be provided for each source as XML XSDcomplexType typically including sensor identification and characteristics.

Special attention shall be given on geo-location data associated to the monitored scene; such set of information shall be regrouped as a XML XSDcomplexType, including GPS latitude and longitude of each item (centre of image and vertex). As required for constructions, buildings, underground structures, etc., means to locate the scene in 3 dimensions in relation to a geo-located reference point shall be provided (scanned drawing or equivalent).

Type of content (video clip, audio or video frame) shall be provided; special contents (e.g. dynamic metadata or events) may be used.

The codec used to encode the relevant content shall be given among:

— H.264;

— JPEG 2000;

— JPEG;

— MPEG-4 Visual;

— Audio codec (see 5.2.4).

Data encoded as mentioned above may be stored by using one of the following container file formats:

— JFIF;

— JP2;

— MP4.

Resolution of the video (XML XSDsimpleType) shall be provided as well in number of pixels in height and width.

NOTE 2 To accommodate legacy systems the AVI, MOV, and WMV widely accepted containers are allowed for existing assets.

The package metadata shall include the data items defined in 5.2.5.

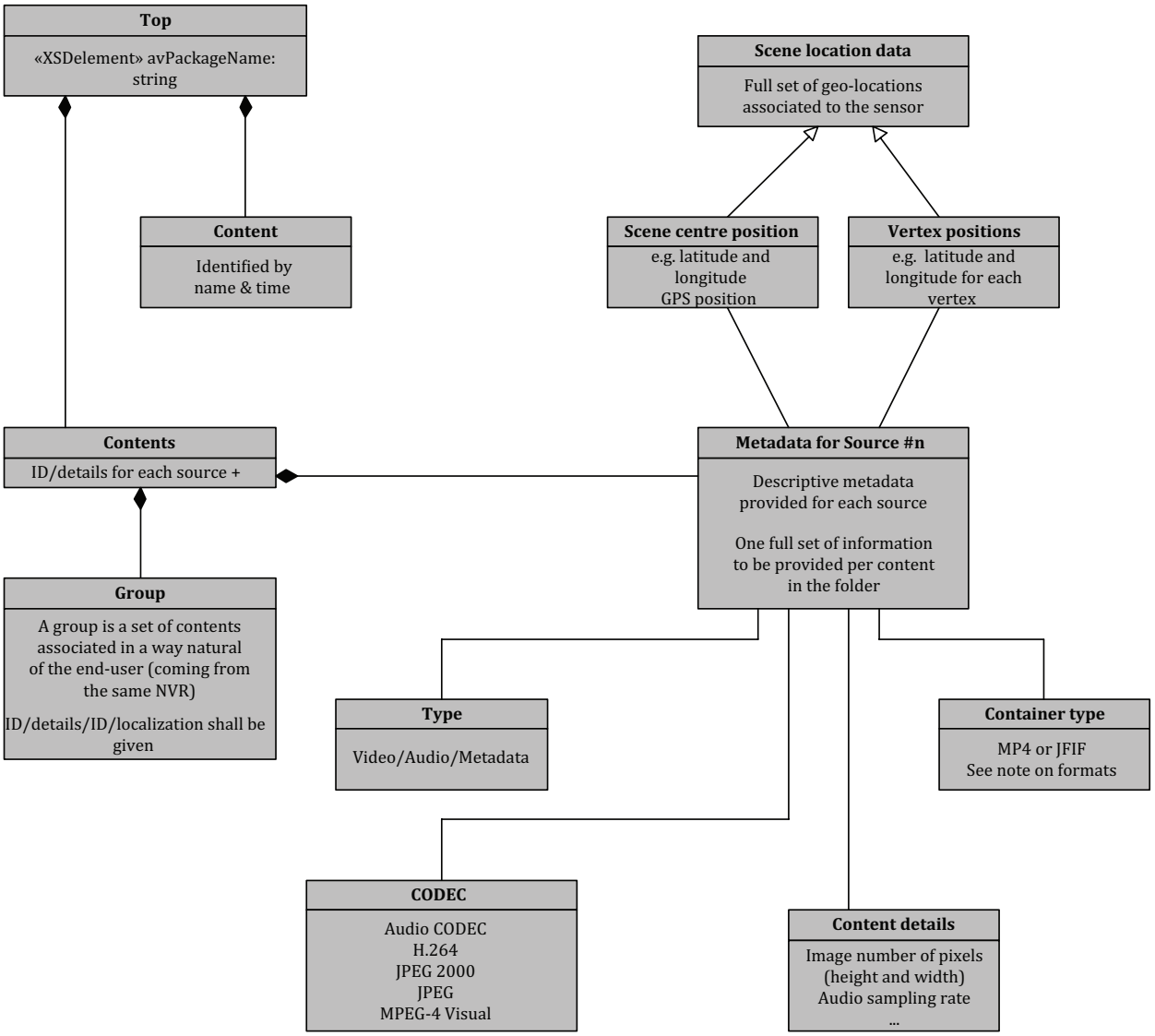

Figure 3 — Structure of the Audio-Video Package XML description and integration in the folder

5.2.1.1 Audio-Video Package XML Descriptor

All the descriptive data items of an Audio-Video Package shall be defined as elements of an XSD extracted from the XML Metadata scheme and namespace defined in IEC 62676-2-3.

Minimum requirements, which rely on correspondence and compliance with maintained dictionaries and practices, are detailed below.

The XML Descriptor of the Audio-Video Package shall be organized as per Figure 4 XML Schema diagram and be stored in DTS Directory and duplicated (optional) in the Description DTS Directory.

As further detailed in 5.2.2.1 two levels of implementation of this International Standard are possible.

To allow the most efficient forensics activities with the individual video (or audio) contents, the Level 2 Systems associate all the XML descriptors specified in the Audio-Video package to the video and audio contents and in case of dynamic metadata, time references allow for correlation with the associated audio and/or video. Such descriptive items, mandatory in Level 2 systems, are specified hereafter.

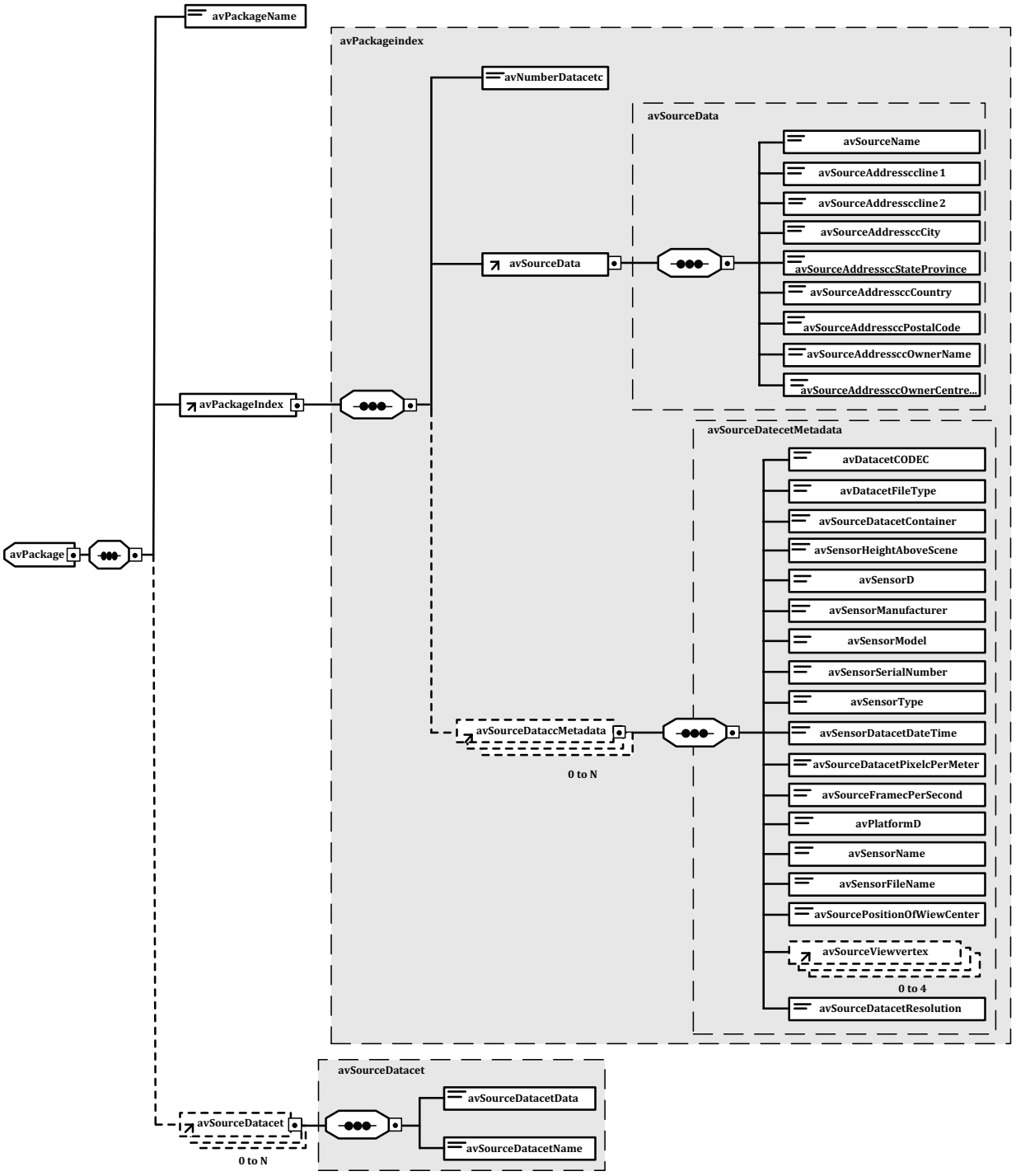

Figure 4 — Arrangement of the XML Descriptor

It is recommended that Level 1 systems also implement to the largest extent these metadata items. Level 1 systems shall nevertheless provide the following minimum metadata:

— Codec name and profile;

— Name of the container;

— Video resolution;

— Video frame rate (in fps);

— Time and date of the record;

— Time and date of the camera.

5.2.2 Requirements for the video source format

5.2.2.1 General

The primary expectation of law enforcement entities is that complying systems will export videos that can be identified and are in a format that can be displayed. This International Standard satisfies this requirement by defining a list of allowed compression schemes (and associated minimum constraints) that are accommodated by the commonly available players. All video-surveillance systems that export data as per this International Standard and compress the video in one of the formats defined in the list below are deemed Level 1 compliant.

To expedite efficiently and without delay complex forensics missions, the requirements go beyond the sole display of the video, and include exploitation of dynamic metadata, accurate and flexible navigation, compatibility with automated video analysis, plug-and-play features, compliance test procedures etc. They require a strict adherence to a video compression scheme and implementation rules, referred as Level 2. These implementation rules include metadata prescriptions defined in 5.2.5.

Level 2 compliant systems are also compliant to Level 1 and it is possible that these two levels are mandated in a same jurisdiction, typically based on implementation dates, level of risk of the monitored infrastructure or number of cameras installed.

5.2.2.2 Requirements for Level 1 Systems

Level 1 systems shall compress video for insertion in the standard export format as per one of the codecs listed below:

— H.264/MPEG4-AVC as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10:2012;

— MPEG-4 Visual as defined in ISO/IEC 14496-2:2004;

— JPEG as defined in Rec. ITU-T T.83 | ISO/IEC 10918-1:1994;

— JPEG 2000 image coding system as per ISO/IEC 15444-1:2004.

Level 1 Systems may supply video in a Level 1 codec utilizing one of the following containers.

— JFIF as defined in ISO/IEC 10918-5 (to be published);

— JP2 as defined in ISO/IEC 15444-1:2004;

— MP4 as defined in ISO/IEC 14496-14:2003and ISO/IEC 14496-15:2010.

NOTE To accommodate legacy systems, the AVI, MOV and WMV widely accepted containers are allowed for existing assets.

Level 1 Systems shall supply video with the following characteristics:

— The resolution (image size), the frame rate and quality level of exported video shall not be less than the one of the recorded video in the system.

5.2.3 Level 2 systems

Level 2 systems shall comply with the following prescriptions:

— The video sources to be integrated in the export format shall be progressive and up to HD. Usage of interlaced and/or SD sources remains however acceptable to cover legacy situations;

— Compression shall be compliant with H.264/MPEG4-AVC as defined in Recommendation ITU-T H.264 | International Standard ISO/IEC 14496-10:2012;

— The profile used shall be either Constrained Baseline, Baseline, Main, or High, and the level shall have a maximum value of 4.0 (all levels below 4.0 are accordingly allowed);

— It shall be possible to associate to each video frame its absolute capture time (with an accuracy better than one video frame and in any case better than 100 ms referred to UTC). Individual data streams only (video, audio and metadata), shall be used. They may be logically regrouped (through the Audio-Video Package XML Descriptors) per sensor or other grouping as shown in Figure 5;

— Each video container shall contain an unlimited number of elementary contiguous time slots. The duration of a time slot shall be comprised between 1 and 600 s. Within a time slot indexes shall allow to point accurately to any specific time.

The overall logical arrangement shall be accordingly as in Figure 5 below, where the DTS and FTS are as per paragraph 5.2.1 definitions.

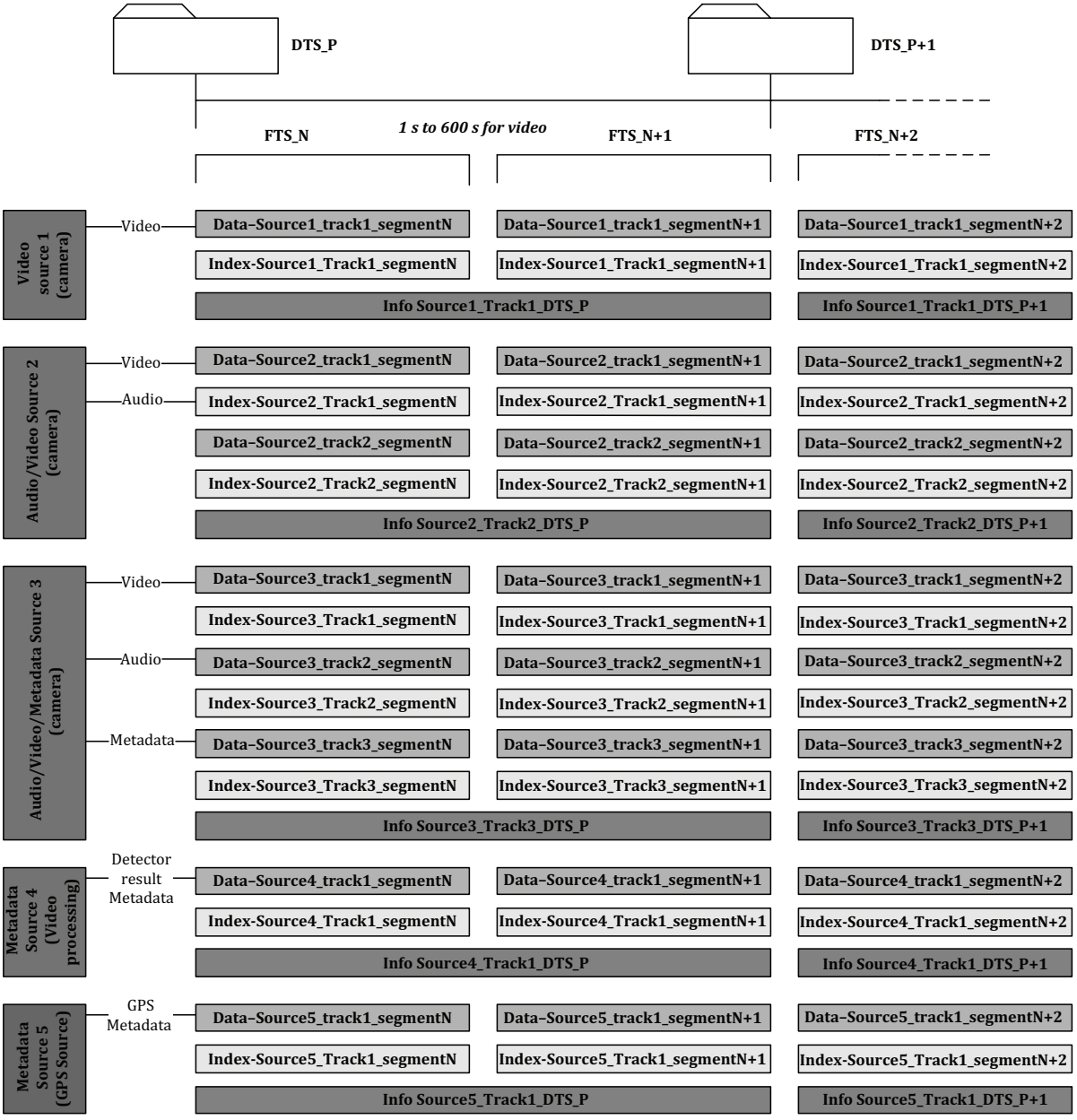

Figure 5 — Arrangement of the descriptive metadata

As several individual data streams from one or more exported files produced by one or more systems require easy and accurate cut and edit, based on capture time, an absolute coarse indexing mechanism shall be implemented, represented as a track in Figure 5.

NOTE 1 Index (fine indexing) is optional.

It is allowed that at the high time resolution level, the formatting at the DTS and the FTS level be performed centrally (typically at a NVR level) in a synchronized manner for all contents, often with a natural sequence (like an "I frame" for video) starting a FTS, whatever the relative phasing between the sensors is. Each individual data stream shall nevertheless carry in addition to this format time index, its capture time index in Coordinated Universal Time (UTC), with a resolution of at least 10 ms. This allows multi-channel players to synchronize videos (and audio) down to the frame level.

This shall be achieved by the implementation for each individual container (elementary data stream during one elementary time slot) of a MPEG-A video surveillance MAF file format as defined in ISO/IEC 23000-10, constituted of one header and of the single sequence of continuous data (a single data chunk as per ISO/IEC 14496-12:2012) corresponding to the elementary time slot. This header shall contain at least the capture time at the beginning of the sequence and an updatable time (which may be the system time at formatting).

This, combined with the descriptive indexing data, shall provide as a minimum for each frame:

— The capture source time;

— The formatting time;

— The type of frame;

— The time index with reference to the beginning of the FTS.

When a content (data source) is dynamic metadata embedding timed information, e.g. scene location for systems involving PTZ cameras or vehicles, this embedded time shall also be the above defined capture source time, ensuring conservation of synchronism between video and associated metadata.

NOTE 2 The compliant video-surveillance systems that produce freeze-frame thumbnail images will export such still images in JPEG Baseline as in ISO/IEC 10918-1/ITU-T T.81.

NOTE 3 Compliant systems may produce videos at levels up to 5.0, but then they become compliant to Level 1 rather than Level 2.

NOTE 4 If jurisprudence in a jurisdiction forbids inter-frame compression, I-only H264 compression can be implemented, while remaining compliant with Level 2.

NOTE 5 All references to MPEG-4 containers in this International Standard are meant as per the above container description.

5.2.4 Audio

When audio is part of the export format, it shall be encoded either as per G.711 Law A as defined in ITU-T/Rec. G.711 or per MPEG4-AAC in Low Complexity Profile (AAC-LC) as defined in ISO/IEC 14496-3:2009. MPEG-4 file format shall be used as per ISO/IEC 14496-14:2003.

5.2.5 Metadata

5.2.5.1 Requirements regarding the metadata items (Level 2 systems)

Metadata shall allow unambiguous definition of each audio-video source or event; this shall include, further to time of occurrence referred to UTC, the absolute geo-location of the four corners of each video frame (scene location), with an accuracy better than 20 % of the smallest height or width of the field of view. This absolute scene location may be obtained through calculations (typically based on PTZ data).

Such descriptive (not changing over the time) unambiguous geo-location information shall be part of the Audio-Video XML Descriptor detailed in 5.2.1.1 and shall rely on the IEC 62676-2-3 metadata format and definitions.

To avoid differences of interpretation, the geodesic metadata definitions are kept in a dictionary in accordance with the requirements described in Annex A. The dictionary structure is as per SMPTE 335M-2001. This dictionary is based on the spatio-temporal subset (Class 7) of the SMPTE metadata dictionary (SMPTE RP210.11-2008); it may be updated over time, in case items are missing for the video-surveillance applications or harmonization with IEC 62676-2-3 becomes necessary.

When the monitored asset is a building, an underground infrastructure or any disposition that cannot rely on the sole GPS coordinates, means to locate the scene in 3 dimensions in relation to a geo-located reference point shall be provided (scanned drawing or equivalent). The mandatory location data (static or dynamic) shall then rely on this absolute reference.

When dynamic metadata are necessary, e.g. for vehicle GPS or PTZ parameters, metadata shall be encoded using the Key-Length-Value (KLV) method as per SMPTE 336M-2007 containing imbedded time reference for time stamping.

When XML metadata need to be used, such as relay of IEC 62676-2-3, XML metadata or results of sophisticated video analytics, XML based descriptors shall be used and embedded in KLV structure.

As a minimum requirement, the format shall incorporate the mandatory metadata allowing source identification and scene location as listed below (in bold in the diagrams of 5.2.5.2 and 5.2.6). Nevertheless processing systems will be built to be resilient to contents which do not contain all the mandatory metadata.

The metadata content can be split into two topics: sensor and event information.

The applicable dictionary and associated KLV encoding are provided in Annex A.

5.2.5.2 Sensor metadata items (Level 2 Systems)

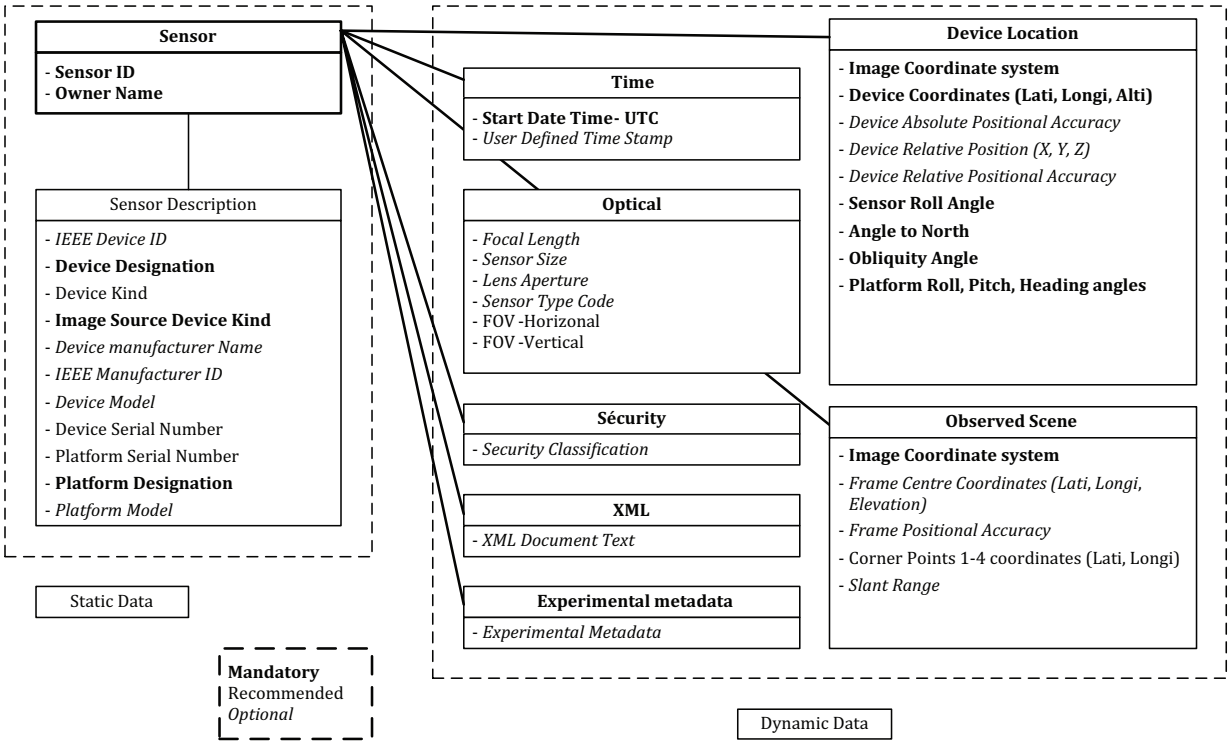

Figure 6 — Sensor metadata items (Level 2 Systems)

The sensor description information concerns the hardware, but also sensor location (including platform position), sensor status, time, optical configuration and security matters. Space is provided to add other information ("Free Item"). Mandatory metadata apply also to audio sensors.

The following metadata items shall be provided:

— Codec name and profile;

— Name of the container;

— Video resolution;

— Video frame rate (in fps);

— Time and date of the record;

— Time and date of the camera.

5.2.6 Event metadata items (Level 2 systems)

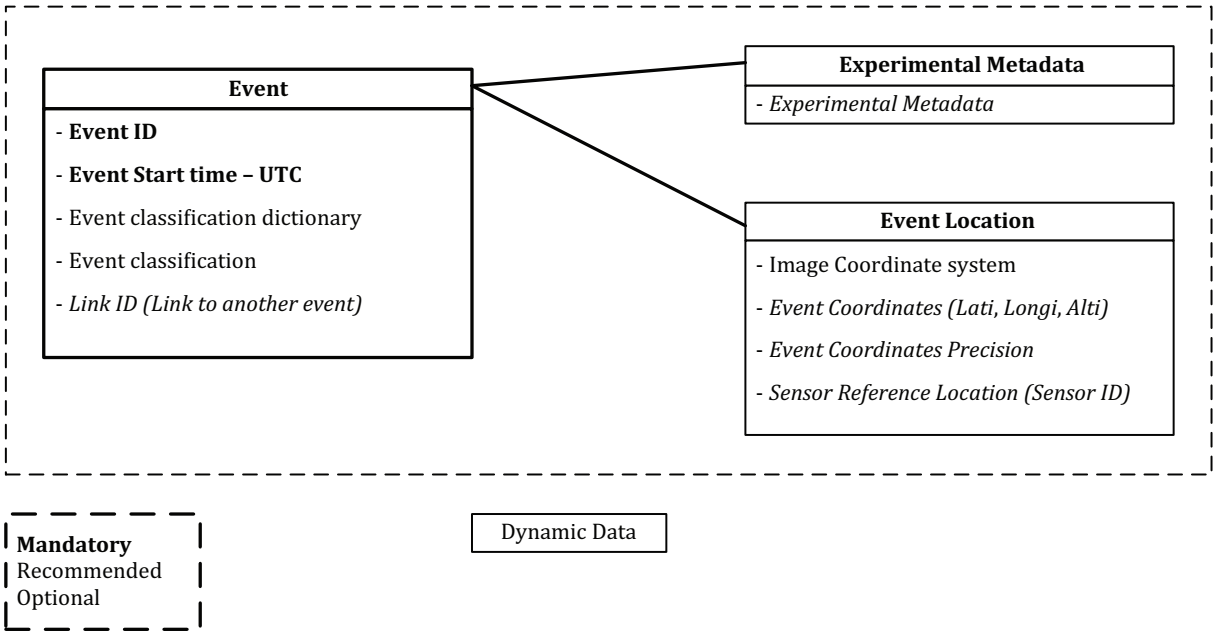

Figure 7 — Event metadata items (Level 2 systems)

This part of the metadata concerns the observed events: it contains time information and event description through the use of specific, domain-related dictionaries. Space is provided to add other information ("Free Item").

Events are by nature dynamic data.

5.3 Data security and integrity

Depending upon local regulations export format may incorporate provisions for mechanisms ensuring that:

— Collected contents have not been corrupted or modified;

— Hierarchical access rights control system can be implemented;

— Access log-ins can be traced.

It is recommended to implement as far as possible data security and integrity specifications published by ISO/IEC JTC 1/SC 27.

5.4 Provisions for privacy

Depending upon local regulations minimum requirements to be implemented, if any, shall at least include:

— Monitoring of the access to the data;

— Mandatory recording time to check that data are properly erased after the prescribed time;

— Minimal masking techniques;

— Training of staff to handle sensitive data.

It is recommended to implement as far as possible privacy specifications published by ISO/IEC JTC 1/SC 27.

Annex A
(normative)

Metadata dictionary

A.1 Encoding rule

Derived from SMPTE 336M, the table in A.2 provides for each item of the metadata dictionary the KLV triplets in the full format, also called Universal Set. Universal Sets use the full KLV Coding Construct throughout.

As per SMPTE 336M, more efficient methods are allowed:

— Global Sets are defined as per Universal Sets, but offer coding efficiency by sharing a common Key header. This coding gain is lossless and every Key can be fully recovered from the data in the Global Set alone.

— Local Sets are defined as per Universal Sets, but offer coding efficiency through the use of short Local Tags whose meaning is defined within the context of the Local Set. Local Sets retain the KLV data construct but require a separate Standard or RP to define the meaning of the Local Tags and to provide a map from the Local Tag value to the Key value (Recommended).

— Variable-length Packs are defined as a further grouping of data items that eliminates the use of Keys and Local Tags for all individual items within the group. Variable-length Packs therefore rely on a Standard or RP, which defines the order of data items within the pack and the UL of each item in the pack.

— Defined-Length Packs are the most efficient (and least flexible) grouping of data items that eliminates the use of both Keys and Local Tags and removes the length for all individual items within the group. Thus Defined-Length packs rely on a Standard or RP which defines both the order of data items, the length of each data item within the pack and the UL of each item in the pack.

A.2 Detailed table

The items of the detailed table are extracts of SMPTE RP210. Not all relevant items have been extracted (e.g. street address, room number or city code). It is assumed in the current version of this International Standard that such information will be carried by XML descriptions.

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Rights						
Owner Name	The name of the person/organization who owns the copyright	M	06.0E.2B.34.01.01.01.01.02.05.01.02.00.00.00.00	ISO 7-Bit Coded Character Set	127 bytes max	
Device Identifiers Unique identifiers for any device used - cameras, microphones etc.						

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Device Designation	Camera ID - Identifies the "house name" of the device used	M	06.0E.2B.34.01.01.01.01.01.01.20.01.00.00.00.00	ISO 7-Bit Coded Character Set	32 chars max	
Device Model	Identifies the device model used	O	06.0E.2B.34.01.01.01.01.01.01.20.03.00.00.00.00	ISO 7-Bit Coded Character Set	32 chars max	
Device Serial Number	Alphanumeric serial number identifying the individual device	R	06.0E.2B.34.01.01.01.01.01.01.20.04.00.00.00.00	ISO 7-Bit Coded Character Set	32 chars max	
IEEE Device ID	Hex number identifying a device by manufacturer and device number	O	06.0E.2B.34.01.01.01.02.01.01.20.05.00.00.00.00	uint 8	6 bytes	
Device Kind	Device Type expressed as a common name - e.g. camera, audio tape recorder, RAM, Hard disk etc.	R	06.0E.2B.34.01.01.01.03.01.01.20.08.00.00.00.00	ISO 7-Bit Coded Character Set	127 bytes max	
Image Source Device Kind	Sensor type	M	06.0E.2B.34.01.01.01.01.04.20.01.02.01.01.00.00	ISO 7-Bit char string	32 chars max	
Platform Identifiers Organisationally given identifiers for platforms, vehicles, or mounts carrying devices or sensors						
Platform Designation	Identifies the generic name of the platform carrying the device used in capturing or generating the essence	M	06.0E.2B.34.01.01.01.03.01.01.21.01.00.00.00.00	ISO 7-Bit Coded Character Set	32 bytes max	
Platform Model	Identifies the platform model carrying the device used in capturing or generating the essence.	O	06.0E.2B.34.01.01.01.03.01.01.21.02.00.00.00.00	ISO 7-Bit Coded Character Set	32 bytes max	
Platform Serial Number	Alphanumeric serial number identifying the individual platform carrying the device used in capturing or generating the essence.	R	06.0E.2B.34.01.01.01.03.01.01.21.03.00.00.00.00	ISO 7-Bit Coded Character Set	32 bytes max	

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Manufacturer Identifiers						
Device Manufacturer Name	The manufacturer or maker of the device	O	06.0E.2B.34.01.01.01.02.01 .0A.01.01.01.00.00.00	ISO 7-Bit Coded Character Set	Variable	
IEEE Manufacturer ID (Binary data are in Big Endian order)	The IEEE registered ID for a particular manufacturer.	O	06.0E.2B.34.01.01.01.02.01 .0A.01.02.00.00.00.00	Unit 24	3 bytes	
Optical Device Parameters Information about the optical device used						
Focal Length	Focal length of the lens at time of capture	O	06.0E.2B.34.01.01.01.02.04.20.02.01. 01.04.00.00	Floating Point	4 bytes	
Focal Length	Focal length of the lens in millimetres at time of collection	O	06.0E.2B.34.01.01.01.03.04.20.02.01. 01.04.01.00	Integer	4 bytes	
Sensor Size	The size of the sensor - e.g. 1/2 inch, 2/3 inch etc.	O	06.0E.2B.34.01.01.01.02.04.20.02.01. 01.05.00.00	ISO 7-Bit Coded Character Set	32 bytes max	
Lens Aperture	Aperture of the lens at the time of collection	O	06.0E.2B.34.01.01.01.02.04.20.02.01. 01.06.00.00	Floating Point	4 bytes	
Sensor Type Code	Code indicating type of sensor that produced the original video content	O	06.0E.2B.34.01.01.01.02.04.20.02.01. 01.07.00.00	ISO 7-Bit Coded Character Set	32 bytes max	
Field of View (FOV-Horizontal)	Sensor Horizontal field of view	R	06.0E.2B.34.01.01.01.02.04.20.02.01. 01.08.00.00	Floating Point	4 bytes	Degrees
Field of View (FOV-Vertical)	Sensor Vertical field of view	R	06.0E.2B.34.01.01.01.07.04.20.02.01.0 1.0A.01.00	Floating Point	4 bytes	Degrees
Positional System Information Information about absolute system in use						

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Image Coordinate system	Identifies the Digital Geographic Information Exchange Standard (DIGEST) geo-referenced coordinate system used at image capture	M	06.0E.2B.34.01.01.01.01.07.01.01.01.00.00.00.00	ISO 7-Bit Coded Character Set	4 bytes	”WG84”
Device Absolute Position (the absolute position of the capturing device)						
Device Longitude	Specifies a sensor’s geographic location in decimal degrees of longitude	M	06.0E.2B.34.01.01.01.03.07.01.02.01.02.06.02.00	Double	8 bytes	Degrees
Device Latitude	Specifies a sensor’s geographic location in decimal degrees of latitude	M	06.0E.2B.34.01.01.01.03.07.01.02.01.02.04.02.00	Double	8 bytes	Degrees
Device Altitude	Altitude of sensor as measured from Mean Sea Level (MSL)	R	06.0E.2B.34.01.01.01.01.07.01.02.01.02.02.00.00	Floating Point	4 bytes	Metres
Device Absolute Positional Accuracy	Accuracy of frame centre coordinates as a Circular Error Probable (CEP) (50 %), (default metres)	O	06.0E.2B.34.01.01.01.01.07.01.02.01.02.01.00.00	Floating Point	4 bytes	Metres
Device Relative Position (the relative position of the essence-capturing device)						
Device Relative Positional Accuracy	Accuracy of frame centre coordinates	O	06.0E.2B.34.01.01.01.01.07.01.02.02.02.01.00.00	Floating Point	4 bytes	Metres
Device Relative Position X	Defined by the X translational position of the camera from a local Datum Absolute Position. Positive values indicate translations in which the camera has physically moved from left to right, (default metres)	O	06.0E.2B.34.01.01.01.01.07.01.02.02.02.02.00.00	Floating Point	4 bytes	Metres

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Device Relative Position Y	Defined by the Y translational position of the camera from a local Datum Absolute Position. Positive values indicate translations in which the camera has physically moved to a higher elevation, (default metres)	O	06.0E.2B.34.01.01.01.01.07.01.02.02.02.03.00.00	Floating Point	4 bytes	Metres
Device Relative Position Z	Defined by the Z translational position of the camera from a local Datum Absolute Position. Positive values shall indicate translations in which the camera has physically moved towards the target, (default metres)	O	06.0E.2B.34.01.01.01.01.07.01.02.02.02.04.00.00	Floating Point	4 bytes	Metres
Distance Measurements						
Slant Range	Distance from the sensor to the centre point on ground of the image (metres)	R	06.0E.2B.34.01.01.01.01.07.01.08.01.01.00.00.00	Floating Point	4 bytes	Metres
Subject Absolute Position (the absolute position of the subject depicted in the essence)						
Frame Centre Latitude	Specifies the video frame centre point geographic location in decimal degrees of latitude. Positive values indicate northern hemisphere; negative values indicate southern hemisphere.	O	06.0E.2B.34.01.01.01.01.07.01.02.01.03.02.00.00	Double	8 bytes	Degrees

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Frame Centre Longitude	Specifies the video frame centre point geographic location in decimal degrees of longitude. Positive values indicate eastern hemisphere; negative values indicate western hemisphere.	O	06.0E.2B.34.01.01.01.01.07.01.02.01.03.04.00.00	Double	8 bytes	Degrees
Frame Centre Elevation	The elevation of field of view frame centre, measured in meters above WGS-84 spheroid	O	06.0E.2B.34.01.01.01.0A.07.01.02.01.03.16.00.00	Floating Point	4 bytes	Metres
Frame Positional Accuracy	Accuracy of frame centre coordinates as a Circular Error Probable (CEP) (50 %), (default metres)		06.0E.2B.34.01.01.01.01.07.01.02.01.03.01.00.00	Floating Point	4 bytes	Metres
Corner Latitude Point 1 (upper left corner) - *corners not corresponding to geographical locations, i.e. above horizon, shall not be included* -	Latitude coordinate of corner 1 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.07.01.00	Double	8 bytes	Degrees
Corner Longitude Point 1 (upper left corner)	Longitude coordinate of corner 1 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.0B.01.00	Double	8 bytes	Degrees
Corner Latitude Point 2 (upper right corner)	Latitude coordinate of corner 2 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.08.01.00	Double	8 bytes	Degrees
Corner Longitude Point 2 (upper right corner)	Longitude coordinate of corner 2 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.0C.01.00	Double	8 bytes	Degrees
Corner Latitude Point 3 (lower right corner)	Latitude coordinate of corner 3 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.09.01.00	Double	8 bytes	Degrees
Corner Longitude Point 3 (lower right corner)	Longitude coordinate of corner 3 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.0D.01.00	Double	8 bytes	Degrees
Corner Latitude Point 4 (lower left corner)	Latitude coordinate of corner 4 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.0A.01.00	Double	8 bytes	Degrees

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Corner Longitude Point 4 (lower left corner)	Longitude coordinate of corner 4 of an image or bounding rectangle	R	06.0E.2B.34.01.01.01.03.07.01.02.01.03.0E.01.00	Double	8 bytes	Degrees
Device Angles Device information regarding angles related to positioning information						
Sensor Roll angle	Specifies the roll angle of the sensor. Expressed in degrees.	M	06.0E.2B.34.01.01.01.01.07.01.10.01.01.00.00.00	Floating Point	4 bytes	Degrees
Angle to North	Angle in degrees from the first row of the image to true north.	M	06.0E.2B.34.01.01.01.01.07.01.10.01.02.00.00.00	Floating Point	4 bytes	Degrees
Obliquity angle	Obliquity angle of image expressed in degrees. The inverse of sensor depression angle.	M	06.0E.2B.34.01.01.01.01.07.01.10.01.03.00.00.00	Floating Point	4 bytes	Degrees
Platform Roll angle	Roll angle of platform expressed in degrees. The Roll of an airborne platform is rotation about its longitudinal (front-to-back) axis; wings level is zero degrees, positive (negative) angles describe a platform orientation with the right wing down(up). Range of values is –180 to +180 degrees	M	06.0E.2B.34.01.01.01.07.07.01.10.01.04.00.00.00	Floating Point	4 bytes	Degrees

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Platform pitch angle (often called pan for a camera)	Pan (or pitch) angle of platform expressed in degrees. The Pitch of an airborne platform describes the angle of its longitudinal (front-to-back) axis makes with the horizontal (i.e. equi-potential gravitational surface); positive (negative) angles describe a nose up (down) orientation. Range of values is −90 to +90 degrees.	M	06.0E.2B.34.01.01.01.07.07.01.10.01.05.00.00.00	Floating Point	4 bytes	Degrees
Platform heading angle (often called tilt for a camera)	Tilt (or heading) angle of platform expressed in degrees. The Heading of an airborne platform is the angle from True North of its longitudinal axis projected onto the horizontal plane. Range of values is 000 to almost 360 degrees; North is 000, East is 090; South is 180, and West is 270.	M	06.0E.2B.34.01.01.01.07.07.01.10.01.06.00.00.00	Floating Point	4 bytes	Degrees
Dates and Times						
Start Date Time - UTC	Absolute time at start of creating the segment, shot, clip, item etc.	M	06.0E.2B.34.01.01.01.01.07.02.01.02.01.01.00.00	ISO 7-Bit Coded Character Set	32 chars max	ISO 8601 short format
User Defined Time Stamp	time, defined as the number of micro-seconds elapsed since midnight Coordinated Universal Time (UTC) of January 1, 1970	O	06.0E.2B.34.01.01.01.04.07.02.01.01.01.05.00.00	Bitwise mapping of 64-bit time-code into 8 bytes, lsb first	8	micro-second
Event Start time - UTC	The absolute beginning date and time of the project, mission, scene, editing event, license, publication etc.	O	06.0E.2B.34.01.01.01.01.07.02.01.02.07.01.00.00	ISO 7-Bit Coded Character Set	32 chars max	ISO 8601 short format
Security						

Data Element Name	Data Element Definition	M = Mandatory R = Recommended O = Optional	Key	Data type	Data length (bytes)	Unit
Security Classification	Marking of the security level or other description of the classification of information	O	06.0E.2B.34.01.01.01.03.02.08.02.01.00.00.00.00	ISO 7-Bit Coded Character Set	14 bytes max	
XML Constructs and Interpretations						
XML Document Text	An XML document as text	O	06.0E.2B.34.01.01.01.05.03.01.02.20.01.00.00.00		Variable	
Experimental metadata						
Experimental metadata	Users may create their own structures consistent with the metadata encoding standard	O	06.0E.2B.34.01.01.01.01 .0F.00.00.00.00.00.00.00		Variable	
Complementary Attributes						
Sensor ID	Unique ID of the sensor	M		ISO 7-Bit Coded Character Set	32 bytes max	
Event ID	Unique ID of the event	M		ISO 7-Bit Coded Character Set	32 bytes max	
Event Classification Dictionary	Name of the Classification Dictionary	R		ISO 7-Bit Coded Character Set	127 bytes max	
Event Classification	Event from the Classification Dictionary	R		ISO 7-Bit Coded Character Set	127 bytes max	

Annex B
(informative)

Requirements concerning usability

B.1 General environment

Where auditory records are being checked the ambient noise levels should be adequately controlled such that essential signals are not masked.

The thermal environment should contribute to the maintenance of operator alertness.

Glare on display screens should be minimised by the use of appropriate artificial light fittings and blinds on windows.

Display screens should not viewed against a backdrop of unscreened windows to avoid disability glare and eye-strain.

Where written records need to be made or consulted, during video record examination, appropriate levels of lighting should be provided.

The requirements of older operators should be taken into account as far as sensitivity to glare and lighting levels.

The interior treatment of the immediate areas, where the video surveillance tasks are to be conducted, should avoid strong hues and visual patterns.

Whenever applicable, ISO 9241 and ISO 11064 will be taken into consideration.

B.2 Furniture layout

The arrangement of furniture should be ergonomically arranged to take account of the primary equipment as well as any use of reference material, writing areas or space for laptops.

Where display screens are mounted in furniture appropriate adjustment should be provided such that operators can easily adjust positions of screens from their seated position.

Where used for extended periods[6)] of time, workstations should comply with display screen equipment.

B.3 Equipment selection

Screen sizes should take account of the likely viewing distances and the range of target sizes operators will be required to detect.

Control panels should cater to both right and left-handed users.

Any critical buttons, such as operating a delete command, should be located such that accidental activation is minimised.

The selection of appropriate controls should take account of the differing characteristics of the controls and the task requirements placed on the operator. Some typical operating characteristics are tabulated below:

6) In Europe, compliance should be with Directives on the use of ‘display screen equipment’.

Table B.1 — Considerations in the selection of control devices

Control/Input	Considerations for Appropriate Use
Cursor Control Keys	Moving cursor in X and Y dimensions
Touch Screen	Moving/holding arm to screen for long periods of time is not required Screen does not have small 'poke points' relative to size of finger tip A low level of resolution is required for positioning Task will not be disrupted by hand temporarily blocking screen Periodic cleaning of screen is provided
Mouse	Adequate space is available for mouse movement over a pad or desktop A low to medium level of resolution is required for positioning Periodic cleaning is provided
Isotonic Joystick (Displacement)	Positioning accuracy is more important than positioning speed
Trackball	Rapid cursor positioning is desirable Limited space is available for installing an input device
Graphics Tablet	A low to medium level of resolution is required
Isometric Joystick (Force)	Precise or continuous control of two or more related dimensions is required

The relationship between any control and the resulting action should be logical and consistent with sound ergonomics practice.

The design of legends on control panels should take account of the conditions under which they will need to be read and likely viewing distances[7)].

The design of any control functions should minimize the potential for repetitive strain injuries.

B.4 Forensic investigation process

Working periods should be arranged such that regular breaks are taken away from the screen[8)]. Self-paced operations are preferred to those where the rate of presentation of images is not controllable by the operator.

Equipment should be arranged in such a manner that it reflects the logical sequences in which it is likely to be used.

Most frequent actions to be performed should involve the least number of movements/actions.

During rest periods investigators should be discouraged from carrying out such screen-based relaxation activities such as computer based games or watching television.

Systems designers should take account of human short-term memory limitations when considering the design of the interfaces associated with meta-data.

Provision should be made for recovery from error.

Training should include the recognition of when fatigue occurs and when to change tasks or take suitable rest breaks.

Intentional variation of tasks and activities should be considered in order to help keep operators alert.

---

7) Legends may need to be viewed under reduced ambient lighting conditions.

8) Typically XXX minutes rest is recommended every YYY.

B.5 Interface design

The presentation of meta-data should adopt ergonomic principles taking account of the way in which this information will be used by operators (dynamic geographical or alphanumerical representation); the presentation of metadata should take into account of the most effective ways of presenting the information including such factors as font sizes, font styles, screen location, backgrounds and language.

The event data should be represented consistently with metadata.

Software design should facilitate the investigators task though its adoption of ergonomic principles.

B.6 Maintenance

Installation of CCTV forensic workstations should take account of the requirements of maintainers to access items of equipment and removing them without excessive physical strain.

Suitable and sufficient access space should be provided around equipment to allow for ease of maintenance.

Adequate lighting should be provided for any *in situ* maintenance tasks.

Bibliography

[1] ISO 9241 (all parts), *Ergonomic requirements for office work with visual display terminals (VDTs)*

[2] ISO 11064 (all parts), *Ergonomic design of control centres*

[3] ISO/IEC 13818-1:2007, *Information technology — Generic Coding of Moving pictures and associated audio information systems*

[4] ISO/IEC 15938-2:2002, *Information technology — Multimedia content description interface — Part 2: Description definition language*

[5] ISO/IEC 15938-3:2002, *Information technology — Multimedia content description interface — Part 3: Visual*

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

[7] ISO 19115:2003, *Geographic information — Metadata*

[8] ISO 19115-2:2009, *Geographic information — Metadata — Part 2: Extensions for imagery and gridded data*

[9] ISO/TS 19130, *Geographic information — Imagery sensor models for geopositioning*

[10] ITU-R BT.1700, *Characteristics of composite video signals for conventional analogue television systems*, 2005

[11] ITU-R BT.601-6, *Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*, January 2007

[12] ITU-R BT.709-5, *Parameter values for the HDTV standards for production and international programme exchange*, April 2002

[13] ITU-R BT.1358: 1998, *Parameters of 625 and 525 line Progressive Scan Television Systems*

[14] ITU-T H.264-1, *Conformance specification for H.264 advanced video coding*, March 2005

[15] SMPTE RP217-2001, *Nonsynchronized Mapping of KLV Packets in MPEG-2 System Streams*

[16] STANAG 4609, *NATO, Digital Motion Imagery Format*

[17] W3C XML Schema Definition Language (XSD) 1.1 (all parts)

ICS 03.100.01

Price based on 29 pages