INTERNATIONAL STANDARD

ISO/IEC 14863

First edition
1996-07-15

Information technology — System-Independent Data Format (SIDF)

Technologies de l'information — Format de données indépendantes du sytème (SIDF)

INTERNATIONAL ISO/IEC

Reference number
ISO/IEC 14863:1996(E)

Contents

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. 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.

International Standard ISO/IEC 14863 was prepared by ECMA (as Standard ECMA-208) and was adopted, under a special "fast-track procedure", by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, in parallel with its approval by national bodies of ISO and IEC.

Annexes A to C form an integral part of this International Standard. Annexes D and E are for information only.

Introduction

The System Independent Data Format Association (SIDF) was formed in early 1993 by a consortium of industries. ECMA TC15 adopted the work of this committee in early 1994 and developed Standard ECMA-208 on the basis of the SIDF specification.

Information technology – System-Independent Data Format (SIDF)

Section 1 - General

1 Scope

This International Standard specifies a logical format for information interchange and secondary data storage. The format provides a System-Independent Data Format (SIDF) for the representation of primary file system information. This information includes, among other things, data, attributes and characteristics. This International Standard specifies

- the organization of the information on target media,
- requirements for originating and receiving systems for the processing of the information.

2 Conformance

2.1 Conformance of Media Volumes

A Volume shall be in conformance with this International Standard if all information recorded on it meets the relevant requirements of sections 2 and 3 for the level of partition claimed.

2.2 Conformance of an originating system

An originating system shall be in conformance with this International Standard if it meets the requirements of 14.2.

2.3 Conformance of a receiving system

A receiving system shall be in conformance with this International Standard if it meets the requirements of 14.3 for the Level of conformance claimed.

3 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 646:1991,	*Information technology -- ISO 7-bit coded character set for information interchange.*
ISO/IEC 2022:1994,	*Information technology -- Character code structure and extension techniques.*
ISO 8859-1:1987,	*Information processing -- 8-bit single-byte coded graphic character sets -- Part 1: Latin alphabet No. 1.*
ISO 8859-2:1987,	*Information processing -- 8-bit single-byte coded graphic character sets -- Part 2: Latin alphabet No. 2.*
ISO 8859-3:1988,	*Information processing -- 8-bit single-byte coded graphic character sets -- Part 3: Latin alphabet No. 3.*
ISO 8859-4:1988,	*Information processing -- 8-bit single-byte coded graphic character sets -- Part 4: Latin alphabet No. 4.*
ISO 9660:1988,	*Information processing -- Volume and file structure of CD-ROM for information interchange*
ISO/IEC 13346-1:1995,	*Information technology -- Volume and file structure of write-once and rewritable media using non-sequential recording for information interchange -- Part 1: General.*
ISO/IEC 13346-2:1995,	*Information technology -- Volume and file structure of write-once and rewritable media using non-sequential recording for information interchange -- Part 2: Volume and boot block recognition.*
ISO/IEC 13346-3:1995,	*Information technology -- Volume and file structure of write-once and rewritable media using non-sequential recording for information interchange -- Part 3: Volume structure.*

ISO/IEC 13346-4:1995, *Information technology -- Volume and file structure of write-once and rewritable media using non-sequential recording for information interchange -- Part 4: File structure.*

ISO/IEC 13346-5:1995, *Information technology -- Volume and file structure of write-once and rewritable media using non-sequential recording for information interchange -- Part 5: Record structure.*

ISO/IEC 9945-1:1996, *Information technology -- Portable Operating System Interface (POSIX) -- Part 1: System Application Program Interface (API) [C Language].*

ISO/IEC 10646-1:1993, *Information technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane.*

ISO/IEC 13800:1996, *Information technology -- Procedure for the Registration of identifiers and attributes for volume and file structure.*

ITU Rec. X.25 (1993), *Interface between data terminal equipment (DTE) and data circuit-terminating equipment (DCE) for terminals operating in the packet mode and connected to public data networks by dedicated circuit.*

4 Definitions

For the purposes of this International Standard, the following definitions apply.

4.1 Blank Space: Either one or more NULL Fields, or a BLANK SPACE Field Table.

4.2 Buffer: A set of adjacent sectors, which are logically related A common use of buffers is to encapsulate subsets of file set data so that it can be treated as a whole.

4.3 byte: A string of eight binary digits operated upon as a unit.

4.4 Byte Sequence: An ordered set of bytes.

4.5 Field: The basic logical grouping of data.

4.6 Field Identifier (FID): A Byte Sequence which identifies a Field.

4.7 Field Table: A set of logically associated Fields.

4.8 File: A collection of data that is stored as a single unit. A File contains a representation of a single primary file system entity, such as a Source directory, file, data base object, or data base.

4.9 file mark: A sector containing a control block used as a delimiter to facilitate positioning, as defined by the relevant standard for recording.

Note 1 - Some media may not support file marks.

4.10 File Set: A set of associated Files, sourced from a single Source, recorded within a Volume Set.

4.11 Interleaved File Sets: Two or more File Sets are interleaved if any part of at least one is recorded between any parts of another.

4.12 Iterated Field Set: A sequence of one or more Fields, which may be repeated within a Field Table.

4.13 operating system: The software which controls the file system of the Source.

4.14 originating system: An information processing system which can create a File Set in a Volume Set for the purpose of data interchange with another system.

4.15 Relative Byte Position: Indicates the byte position in a number.

4.16 receiving system: An information processing system which can process a File Set of a Volume Set which has been created by another system for the purpose of data interchange.

4.17 Sector: The data field of the smallest addressable part of the medium that can be accessed independently of other addressable parts of the medium.

Note 2 - This Sector is different from the physical sectors of a physical track.

4.18 Sector Number: The ordinal sequence number of a Sector within a Volume, starting with 0 (0, 1, 2, 3, ..., n).

4.19 Source: The primary system (e.g. file system, data base) which is the source of the Files that are recorded as a File Set for secondary storage. The name of the Source is specified for the File Set, and is not included in the Source path information of the individual Files.

4.20 standard for recording: A standard that specifies the recording method and the addressing method for the information recorded on a medium.

4.21 Volume: A set of Sectors which resides within one physical partition on a secondary storage medium.

4.22 Volume Set: One Volume, or multiple logically related Volumes.

4.23 Volume Set Sequence Number: The ordinal sequence number of each Volume within a Volume Set, starting with 1 (1, 2, 3, ..., n).

5 General conventions and notations

5.1 Representation of numbers

- Numbers in decimal notation are shown as decimal digits.
- Numbers in hexadecimal notation are shown as a sequence of hexadecimal digits preceded by #.
- The setting of bits is denoted by ZERO or ONE.
- Numbers in binary notation and bit combinations are represented by sequences of ZEROs or ONEs.
- Bit combinations are shown with the most significant bit to the left.
- Bits are indicated with the small letter b.
- Bytes are indicated with the capital letter B.

5.2 Names

The names of formally defined entities, such as a Volume, are shown with a capital initial.

The name of Fields, such as VOLUME HEADER, are written in capitals.

5.3 Structures

Data structures are shown in tables indicating the Relative Byte Position (RBP), purpose and size of the entities that comprise the structure.

5.4 Fields

Fields are denoted by names shown in capital letters.

5.5 Field Identifiers (FIDs)

A Byte Sequence in hexadecimal notation uniquely identifying a Field.

5.6 Field Tables

The name of a Field Table is that of its first Field, it is shown with capital initials. Field Tables list each Field with its name, its FID and a brief description of the data. The normative definition of each Field is specified in annex C. The Data Description reads "empty" if no Data part is included in the Field. Iterated Field Sets within a Field Table are shown indented, with a bar alongside. Nested Iterated Field Sets are shown with each inner level of nesting further indented than the outer level(s).

5.7 Headers

The term "File Set (Continuation) Header" is used to express: File Set Header and/or File Set Continuation Header. The term "File Record (Continuation) Header" is used to express: File Record Header and/or File Record Continuation Header.

5.8 Indices

The term "Volume (Sub)Index" is used to express: Volume Index and/or Volume Subindex. The term "File Set (Sub)Index" is used to express: File Set Index and/or File Set Subindex.

5.9 Structure description schema

Requirements for recording the different elements of the format are summarized by means of a structure description.

A structure shall be a sequence of terms. A term shall be either:

- a name, enclosed by [], followed by a definition enclosed by { }, or
- a name, enclosed by <>, of a term defined in the structure or of a descriptor defined elsewhere in this International Standard in the text associated with the structure.

Within a definition enclosed by { }, the Or Operator and Repetition Operators may be applied to the terms therein.

The Or Operator is represented by the character VERTICAL LINE: |. It is placed between the terms to which it applies and indicates that either term, but not both simultaneously, is valid at that point in the definition.

There are three Repetition Operators, which shall be interpreted as shown in figure 1, where *n* and *m* are decimal digits.

Repetition Operator	Interpretation
n	*n* occurrences of the preceding term
n+	*n* or more occurrences of the preceding term
n+m	*n* to *m* occurrences of the preceding term

Figure 1 - Repetition Operators

The Repetition Operators shall apply only to the preceding term.

The operators are listed in increasing order of precedence:

| repetition operator [ ] ( )

Higher precedence operators shall be applied before lower precedence operators.

As an example, the schema shown in figure 2 specifies that the entity "Concert" shall be organized as zero or one Introduction, followed by one or more Pieces, an Intermission, one or more Pieces, and zero or more Encores. A Piece consists either of one Song or three to five Movements.

```
[Concert] {
            <Introduction> 0+1
            [Piece] {
                      <Song>
                      | <Movement> 3+5
                     } 1+
            <Intermission>
            <Piece> 1+
            <Encore> 0+
           }
```

Figure 2 - Example of the syntaxof the structure description schema

Section 2 - Requirements for the media

6 Specific conventions and notations for the recorded format

6.1 Recording of numbers

- In each Field the field data is recorded so that the least significant byte (denoted byte B0) is recorded first. Within each byte the least significant bit is denoted bit b0, the most significant bit is denoted b7, and is recorded first. This order of recording also applies to the data of the cyclic redundancy code (CRC) and to its output (CRC bits).

– Negative values are recorded in TWO's complement notation.

6.2 Character sets and coding

Except as specified in this clause and in 6.13, the characters used in the Fields specified by this International Standard shall be coded according to the International Reference Version (IRV) of ISO/IEC 646. The CS4 character set described in 6.8 shall be the default character set, and shall be used if no other is explicitly specified. The specification of the characters allowed in these Fields and the method of recording shall be recorded as a CHAR SPEC Field (see 6.3).

Note 3 - Support for a variety of character ses is a requirement of this International Standard. Ideally, there should be only one character standard used. In practice, several standards, including ISO/IEC 646, ISO/IEC 2022, ISO 8859 and ISO/IEC 10646-1 are used. This International Standard accommodates current practice by specifying several character sets and providing a mechanism for specifying other character sets.

As an example, CS2 (see 6.6) uses the IRV of ISO/IEC 646 as the basic character set but restricts Fields containing characters to a widely usable subset of this character set.

6.3 Character set specification (CHAR SPEC)

The data part of the CHAR SPEC Field consists of two parts, Character Set Type and Character Set Information, recorded as a sequence of bytes. The Character Set Type shall be recorded as the first byte of the sequence; the Character Set Information shall consist of any remaining bytes.

The set of characters allowed in certain Fields and/or Field Tables shall be specified by a one-byte Character Set Type, and additional Character Set Information appended immediately after as needed.

6.3.1 Character Set Type

The Character Set Type identifies the allowed set of characters as shown in figure 3.

Type	Allowed characters
0	The CS0 coded character set (see 6.4).
1	The CS1 coded character set (see 6.5).
2	The CS2 coded character set (see 6.6).
3	The CS3 coded character set (see 6.7).
4	The CS4 coded character set (see 6.8).
5	The CS5 coded character set (see 6.9).
6	The CS6 coded character set (see 6.10).
7	The CS7 coded character set (see 6.11).
8	The CS8 coded character set (see 6.12).
9 to 255	Reserved for future standardization

Figure 3 - Sets of allowed characters

6.3.2 Character Set Information

The interpretation of the contents of this field, if any, is specified by the value of the Character Set Type as described in 6.4 to 6.12.

6.4 CS0 character set

The CS0 character set shall be subject to agreement between the originator and recipient of the medium.

The character set is undefined for interchange. However, the agreed character set can be identified in the Character Set Information, if so desired.

6.5 CS1 character set

The CS1 character set shall be the graphic characters of the character sets specified by the Character Set Information Field.

The Character Set Information Field shall specify one or more escape sequences, to be used in an 8-bit environment according to ISO/IEC 2022 that designate and implicitly invoke graphic character sets specified in ISO/IEC 10646-1. These sequences shall be recorded contiguously from the start of the Field.

6.6 CS2 character set

The CS2 character set shall be the 38 characters in positions 02/14, 03/00 to 03/09, 04/01 to 05/10, and 05/15 of the IRV of ISO/IEC 646.

Note 4 - These characters are: FULL STOP, DIGITs, LATIN CAPITAL LETTERs and LOW LINE.

6.7 CS3 character set

The CS3 character set shall be the 65 characters in positions 02/13 to 02/14, 03/00 to 03/09, 04/01 to 05/10, 05/15, and 06/01 to 07/10 of the IRV of ISO/IEC 646.

Note 5 - These characters are: HYPHEN-MINUS, FULL STOP, DIGITs, LATIN CAPITAL LETTERs, LATIN SMALL LETTERs and LOW LINE.

6.8 CS4 character set

The CS4 character set shall be the 95 characters in positions 02/00 to 07/14 of the IRV of ISO/IEC 646.

6.9 CS5 character set

The CS5 character set shall be the 191 characters in positions 02/00 to 07/14 and 10/00 to 15/15 of Latin Alphabet No. 1 in ISO 8859-1.

6.10 CS6 character set

The CS6 character set shall be the graphic characters of the character sets specified by the Character Set Information Field.

The Character Set Information field shall specify one or more escape sequences according to ISO/IEC 2022 that designate and implicitly invoke the graphic character sets to be used in an 8-bit environment according to ISO/IEC 2022 or ISO/IEC 10646-1. These sequences shall be recorded contiguously from the start of the Field.

6.11 CS7 character set

The CS7 character set shall be the graphic characters of the character sets specified by the Character Set Information Field and code extension characters (see 6.11.1).

The Character Set Information field shall specify one or more escape sequences according to ISO/IEC 2022 that designate and implicitly invoke the graphic character sets to be used in an 8-bit environment according to ISO/IEC 2022 or ISO/IEC 10646-1. These sequences shall be recorded contiguously from the start of the field.

6.11.1 Code extension characters

Characters specified by the CS7 Character Set may include one or more of the following, referred to as code extension characters, to allow alternative character sets to be recorded.

- Escape sequences according to ISO/IEC 2022 or ISO/IEC 10646-1
- Shift functions according to ISO/IEC 2022

6.12 CS8 character set

The CS8 character set shall be the 53 characters in positions 02/01, 02/03 to 02/09, 02/13 to 02/14, 03/00 to 03/09, 04/00 to 05/10, 05/14 to 06/00, 07/11 and 07/13 to 07/14 of the IRV of ISO/IEC 646.

Note 6 - These characters are: EXCLAMATION MARK, NUMBER SIGN, DOLLAR SIGN, PERCENT SIGN, AMPERSAND, APOSTROPHE, LEFT PARENTHESIS, RIGHT PARENTHESIS, HYPHEN-MINUS, FULL STOP, DIGITs, LATIN CAPITAL LETTERs, CIRCUMFLEX ACCENT, LOW LINE, GRAVE ACCENT, LEFT CURLY BRACKET, RIGHT CURLY BRACKET, TILDE.

6.13 Source Name Space specification

The Source or originating system shall indicate the properties of path name information that is associated with a Source file or similar entity by the Field NAME SPACE. In some contexts other Fields may also use the same Name Space information, for example RESOURCE NAME SPACE in the File Set Header. The Name Space specifies the following characteristics of the file name:

- **Path element character set restrictions**: The characters used in the Source file path name shall be contained within this set.
- **Path element length restrictions**: Each element of the path name shall consist of at least one character, and shall not be longer than the number of characters specified.
- **Total path length restriction**: The entire path shall not be longer than the number of characters specified.
- **Path ordering (parent-first or child-first)**: For complete paths in hierarchical file systems, if the parent-first path ordering is used, the parent node closest to the root shall be listed first, followed by its immediate child, and proceeding so forth to the path element furthest from the root. If the child-first path ordering is used, the node furthest from the root shall be listed first, followed by its immediate parent, and proceeding so forth to the path element nearest to the root.
- **Case-sensitive**: A CAPITAL LETTER is considered different from the corresponding SMALL LETTER.
- **Case retention**: A CAPITAL LETTER and the corresponding SMALL LETTER are stored and displayed distinctly.
- **Separators**: In the full path name, the separator(s) between any two elements of the path name.
- **Special requirements**: Any other characteristics not reflected above. For example, whether restrictions apply to the character FULL STOP (#2E), regarding its position or frequency of use in a path element name.

6.13.1 Name Space

Name Spaces shall be identified by an unsigned hexadecimal number. The Name Spaces specified by this International Standard are shown in figure 4.

Identifier	Name Space
#00	The NS0 Name Space (see 6.14).
#01	The NS1 Name Space (see 6.15).
#02	The NS2 Name Space (see 6.16).
#03	The NS3 Name Space (see 6.17).
#04	The NS4 Name Space (see 6.18).
#05	The NS5 Name Space (see 6.19).
#FFFFFFFC	The NSFC Name Space (see 6.20).
#FFFFFFFE	The NSFE Name Space (see 6.21).

Figure 4 - Defined Name Spaces

6.14 NS0 Name Space

Path element character set restrictions: Characters of the NS0 Name Space path elements shall be members of the CS8 Character Set.

Path element length restrictions: A path element shall contain at most 12 characters. A path element shall not contain more than one FULL STOP. A path element shall contain at most 8 characters before a FULL STOP and at most 3 characters after a FULL STOP. If a path element does not contain a FULL STOP, that path element shall contain at most 8 characters.

Total path length restriction: No total path length restrictions shall apply.

Path ordering: The parent path element(s) of a complete (i.e. not relative) path shall be recorded before the child element.

Case-sensitive: FALSE.

Case retention: FALSE.

Separators: The separator between the first two elements of a complete path shall be COLON. The separator between all subsequent elements of the path shall be SOLIDUS.

Special requirements: A path element shall not contain more than one FULL STOP character. The first element of a complete path shall be a Source volume Name.

Note 7 - This Name Space may be associated with file systems of many personal computers. In some implementations the total path length is restricted to 256 characters, and the use of longer paths may yield unpredictable results. In some implementation the characters represented by #21, #23 to 29, #2D, #2E, #7D and #7E may not be supported, and their use may yield unpredictable results.

6.15 NS1 Name Space

Path element character set restrictions: Bytes #00 and #3A are not allowed for the NS1 Name Space path elements.

Path element length restriction: Each path element shall contain at most 32 characters.

Total path length restrictions: No total path length restrictions shall apply.

Path ordering: The parent path element(s) of a complete path shall be recorded before the child element.

Case-sensitive: TRUE.

Case retention: TRUE.

Separators: The separator between the first two elements of a complete path shall be two adjacent COLONs. The separator between all subsequent elements of the path shall be COLON.

Special requirements: The first element of a complete path shall be a Source volume Name.

6.16 NS2 Name Space

Path element character set restrictions: Characters of the NS2 Name Space path elements shall be represented by any single byte, except #00, #2F, #3A.

Path element length restriction: Each path element shall contain at most 300 characters.

Total path length restrictions: No total path length restrictions shall apply.

Path ordering: The parent path element(s) of a complete path shall be recorded before the child element.

Case-sensitive: TRUE.

Case retention: TRUE.

Separators: The separator between the first two elements of a fully qualified path shall be COLON. The separator between all subsequent elements of the path shall be SOLIDUS.

Special requirements: The first element of a complete path shall be a Source volume Name.

6.17 NS3 Name Space

Path element character set restrictions: Characters of the NS3 Name Space path elements shall be represented by any single byte within one of the following ranges: #01 to #2E, #30 to #39, #3B to #7E.

Path element length restriction: Each path element shall contain at most 255 characters.

Total path length restrictions: No total path length restrictions shall apply.

Path ordering: The parent path element(s) of a complete path shall be recorded before the child element.

Case-sensitive: TRUE.

Case retention: TRUE.

Separators: The separator between the first two elements of a complete path shall be COLON. The separator between all subsequent elements of the path shall be SOLIDUS.

Special requirements: The first element of a complete path shall be a Source volume Name.

6.18 NS4 Name Space

Path element character set restrictions: Characters of the NS4 Name Space path elements shall be represented by any single byte within one of the following ranges: #21 to #2E, #30 to #5B, #5D to #FF.

Path element length restriction: Each path element shall contain at most 300 characters.

Total path length restrictions: No total path length restrictions shall apply.

Path ordering: The parent path element(s) of a complete path shall be recorded before the child element.

Case-sensitive: FALSE.

Case retention: TRUE.

Separators: The separator between the first two elements of a complete path shall be COLON. The separator between all subsequent elements of the path shall be SOLIDUS.

Special requirements: The first element of a complete path shall be a Source volume Name.

6.19 NS5 Name Space

Path element character set restrictions: Characters of the NS5 Name Space path elements shall be represented by any single byte within one of the following ranges: #30 to #39, #40 to #5A, #5E to #60, #7B, #7D to #FF.

Path element length restriction: Each path element shall contain at most 256 characters.

Total path length restrictions: No total path length restrictions shall apply.

Path ordering: The parent path element(s) of a complete path shall be recorded before the child element.

Case-sensitive: FALSE.

Case retention: TRUE.

Separators: The separator between the first two elements of a complete path shall be COLON. The separator between all subsequent elements of the path shall be SOLIDUS.

Special requirements: The first element of a complete path shall be a Source volume Name.

Note 8 - In some implementations the total path length is restricted to 256 characters, and use of longer paths may yield unpredictable results.

6.20 NSFC Name Space

Path element character set restrictions: Characters of the NSFC Name Space path elements shall be represented by any single byte within one of the following ranges: #0 to #2A, #2C to #2D, #2F to #3C, #3E to #5B, #5D to #FF.

Path element length restriction: Each path element shall contain at most 256 characters.

Total path length restrictions: The total path shall contain at most 256 characters.

Path ordering: The child element of a complete path shall be recorded before the parent path element(s).

Case-sensitive: FALSE.

Case retention: TRUE.

Separators: The separator between all path elements shall be FULL STOP.

6.21 NSFE Name Space

Name Space NSFE shall be used to indicate that the characteristics associated with the Name Space shall be determined by the Source, and are not specified by this International Standard.

6.22 Strings

A string shall be recorded as a sequence of zero or more non-NULL characters, followed by a NULL character.

6.23 Resynchronization Pattern

The Resynchronization Pattern shall be the two-byte Byte Sequence #A55A.

7 Timestamp

A Timestamp specifies a date and time recorded in the format shown in figure 5. If the Year field has the value 0 the Timestamp shall be ignored; all other Fields in Timestamp shall be set to all ZEROs. Timestamps are established by the originating system and/or the Source. The accuracy of Timestamps is not specified by this International Standard. If a Timestamp is recorded in a Field the Data part of which is longer than 12 bytes, the Timestamp shall be recorded first, and all subsequent bytes shall be recorded as NULL(s).