INTERNATIONAL STANDARD

ISO/IEC 8211

Second edition
1994-10-01

Information technology — Specification for a data descriptive file for information interchange

Technologies de l'information — Spécifications pour fichier de données descriptif pour l'échange d'information

INTERNATIONAL ISO/IEC

Reference number
ISO/IEC 8211:1994(E)

Contents

List of figures

List of tables

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 the 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 8211 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 21, *Open systems interconnenction, data management and open distributed processing*.

This second edition cancels and replaces the first edition (ISO 8211:1985), which has been technically revised.

The substantive changes made to produce this edition of ISO/IEC 8211 are the following additions:

1. Binary forms for numeric values.
2. Binary leaders and directories.
3. Support for ISO/IEC 10646.
4. Definition of FTAM unstructured and structured document types.
5. Concatenated regular structures.
6. Recursive tree description.
7. A human-readable, alternate form of data field description.

The second edition is backwards compatible with the first edition.

Annex A forms an integral part of this International Standard. Annexes B to H are for information only.

Introduction

This International Standard has been produced in response to an identified need for a mechanism to allow data structures to be easily moved from one computer system to another, independent of architecture. Data structures required to be interchanged can vary significantly in complexity and size, and a common method to accomplish these interchanges is desirable. It is also desirable that any medium such as a communication line, a magnetic tape, a disk pack, a flexible disk etc., should be able to be used for the physical interchange, and that all information necessary to successfully recreate the structure in the target system should be contained within the information transported on the medium.

To meet these needs this International Standard specifies medium-independent and system-independent file and data record formats for the interchange of information between computer systems. This International Standard is intended for use with physical recorded media as well as with communications media. The contents in the user data structure can be of any internationally recognized character set and coding and are interchanged in a transparent fashion. The intermediate structure through which the information passes is designed for interchange purposes. It can also be used for some forms of general processing and is amenable to direct access methods on high volume, direct access interchange media.

This International Standard is a concrete transfer syntax and encoding standard and provides a tool for the description of files containing user data but does not specify the content or order of user data fields or user data records. It does specify a comprehensive generic form for such records and fields which can accommodate a wide variety of user needs for both simple and complex user data. An application must design its own instance of a conforming interchange file and all conforming files, both data and data description, will be processable to the field or subfield level by the same software. A user must, of course, complete the interface to their own application system.

The approach used is to define an interchange format into which most information structures and their content can be transformed without loss of information, and from which the original structure and content can be retrieved. The interchange format is suitable both for recording on physical media and transport through a communication system.

The data structures supported by the interchange format are elementary data, vectors, arrays and hierarchies. The file structures that can be transformed into the interchange format include sequential, hierarchical and relational. Network structures are not directly supported and additional pre-processing and post-processing are necessary in this case to preserve logical linkages.

This International Standard is media independent. It assumes, at a minimum, that the supporting transport system can process fixed length octet strings. It requires a computer processing capability to map the user file or database data to the interchange file. This mapping function has to provide the necessary data and structure conversions. The parameters required to define the selection and conversion of these data items and structures into the formats specified by this International Standard are outside the scope of the standard. This International Standard requires the use of a basic character set based on ISO 8859-1 and ISO/IEC 6429 in control fields and permits the use of additional character sets in user data fields. This International Standard provides for three interchange levels from which the users may choose based on the complexity of their data structures. The first interchange level supports multiple fields containing simple, unstructured character strings. The second level supports the first level and multiple fields containing structured user data comprising a variety of data types. The third level supports the second level and hierarchical data structures.

The experience of implementing ISO/IEC 8211 for a variety of applications revealed the need for the changes introduced in this version. Many of the changes give ISO/IEC 8211 increased versatility and more effective interchange capabilities. Many other changes were made to improve clarity and user acceptance. Technical changes in the standardized interchange supporting this International Standard and changes in the organizational responsibility for this International Standard have led to other extensions. This version provides the user with an improved interchange tool in keeping with the user's increasing needs and well integrated into the OSI environment.

The retrieval of archived files may require the use of computer systems which are different from the original archiving systems. The operational problems are identical to those involved in the transport of files between computers at different sites and this International Standard provides a facility for this application.

ISO/IEC 8211 is based on ISO 2709 having the same record structure but different data description components. ISO 2709 based systems for file transfer and random file access had been in use since at least 1970 and their use is now extensive. The nomenclature of ISO/IEC 8211 conforms to its predecessor, ISO 2709. Its hierarchical, logical data constructs are files, variable-length records, variable-length fields and subfields. In several programming languages, the equivalent of an ISO/IEC 8211 field is a record and specific applications will transport their records as ISO/IEC 8211 fields with related records aggregated into ISO/IEC 8211 records.

Organization of the Standard

The contents of this standard are organized as follows:

1) Clause 5 describes the specifications of the contents of leaders, directories and field areas common to all logical records and necessary to the import of logical records and complete fields.
2) Clause 6 describes the specification of the data description necessary to import user data at the subfield level. Subclause 6.1 specifies further subfields in the DDR leader which contain information pertinent to data description.
3) Clause 7 describes the use of extended character sets.
4) Annex A describes the FTAM registrations.
5) Annex B describes a methodology for specifying ISO/IEC 8211 file design and data descriptions.
6) Annexes C through H provide tutorial information on the methodology. The reader may wish to read Annex C prior to studying Clause 5 and Annex D prior to studying Clause 6.

This page intentionally left blank

Information technology – Specification for a data descriptive file for information interchange

1 Scope

This International Standard specifies an interchange format to facilitate the moving of files or parts of files containing data records between computer systems. The interchange format is not intended as a record format for the indigenous files of any specific system but may be used for this purpose. The standard defines a generalized structure which can be used to transmit, between systems, files or records containing a wide variety of data types and data structures. It specifies the means for the description of the contents of data records but does not specify their application semantics although these semantics can be included as a part of the transmission. The interchange format may also be used to transport individual records, individual data fields or individual subfields with their description.

This International Standard specifies:

- media-independent file and data record descriptions for information interchange;
- the description of data elements, vectors, arrays and hierarchies containing character strings, bit strings and numeric forms;
- a data descriptive file composed of a data descriptive record and companion data records that enable interchange to occur with minimal specific external description;
- the data descriptive record that describes the characteristics of each data field within the companion data records;
- three levels of complexity of file and record structure;
- FTAM unstructured and structured document types

2 Normative references

The following standards contain provisions which, through reference in the text, constitute provisions of this International Standard. At the time of its 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 2022:1986, *Information processing - ISO 7-bit and 8-bit coded character sets - Code extension techniques*

ISO 6093:1985, *Information processing - Representation of numerical values in character strings for information interchange*

ISO/IEC 6429:1992, *Information technology - Control functions for coded character sets*

ISO 8571-1:1988, *Information processing systems - Open Systems Interconnection - File Transfer, Access and Management - Part 1: General Introduction*

ISO/IEC 8824:1990, *Information technology - Open Systems Interconnection - Specification of Abstract Syntax Notation One (ASN.1)*

ISO 8859-1:1987, *Information processing - 8-bit single-byte coded graphic character sets - Part 1: Latin alphabet No. 1*

ISO/IEC 9834-2:—1993, *Information technology - Open Systems Interconnection - Procedures for operation of OSI Registration Authorities: Part 2: Registration procedures for OSI document types*

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

IEC 559:1989, *Binary floating point arithmetic for microprocessor systems* (also ANSI/IEEE 754:1985(R1991))

The following document is also relevant to this International Standard:

ISO International register of coded character sets to be used with escape sequences

3 Conformance

Interchange files conform to this International Standard when all of the data descriptive records and data records conform to the specifications of this International Standard. A statement of conformance shall specify the version number and interchange level to which the contents of files conform.

This International Standard does not specify requirements for processing and implementation, therefore such processing and implementation cannot itself conform to this International Standard.

4 Definitions and abbreviations

4.1 Definitions

For the purposes of this International Standard the following definitions apply:

4.1.1 alphanumeric character: A character occurring in columns 2 to 7 inclusive (except position (7/15)) of the Basic Character Set or the corresponding characters of ISO/IEC 10646 Table 1 - Row 00: Basic Latin. These characters correspond to those of ISO/IEC 646 IRV.

NOTE 1 - The characters specified in this International Standard are represented by their position (column/row) in the coded character set table or by their acronym or name, e. g., ESC, SPACE and DIGIT ZERO or "0". Alphanumeric characters which are explicitly specified in a control field are enclosed in double quotes, e.g., "1". The special ISO/IEC 8211 delimiters unit terminator(1/15) and field terminator(1/14) are represented by UT and FT respectively.

4.1.2 array: A data structure of two or more dimensions.

4.1.3 array descriptor: A Cartesian label or numeric array descriptor which provides a description of the dimension and extents of an array.

4.1.4 base address of data: A data element the value of which is equal to the octet count up to but not including the first octet of the first data field following the field terminator of the directory, where the specified origin (0) is the first octet of the leader.

4.1.5 basic character set; BCS: A character set comprising: a) ISO 8859-1 including SPACE as the G0 and G1 sets and b) ISO/IEC 6429 as the C0 and C1 sets.

NOTE 2 - This set is compatible with row 00 of the ISO/IEC 10646 Basic Multilingual Plane and should greatly reduce any dependence on national variant character sets. However there currently are differences in the allowed uses of the C1 set between ISO/IEC 6429 and ISO/IEC 10646.

4.1.6 binary form: A binary form is a sequence of one or more octets having a specified format and meaning. The octets of a binary form may be recorded in order of decreasing significance with the most significant octet first (MSOF) or the reverse order, least significant octet first (LSOF). The bits within an octet are always recorded in order of decreasing significance.

4.1.7 binary tree : A rooted tree in which each node has zero subtrees or, at most, two subtrees known as the left subtree and the right subtree. In the absence of either the left or right subtree, the remaining subtree retains its identity as a left or right subtree. See annex F.

4.1.8 bit field: A data field comprising only binary digits. When it is necessary to complete an octet, it is filled on the right with binary zeros. See also **character mode bit string**.

4.1.9 Cartesian label: An array of identifiers formed by the Cartesian product of the elements of two (or more) vector labels. The array elements have the same order as the elements of the direct product such that if a and b are the vector labels, where a = (a(1) ..., a(*n*)) and b = (b(1) ..., b(*m*)), then the Cartesian label, a*b = (a(1)b(1), a(1)b(2), ..., a(1)b(*m*), ..., a(*n*)b(*m*)), where a(*i*)b(*j*) is a concatenation of a(*i*) and b(*j*) which forms an identifier of the *j,i*-th element of a corresponding data array (i.e., the "a"s are the row labels and the "b"s are the column labels). The Cartesian label is the most general form of a label and contains a vector label and a single label as special cases. The expansion of a higher order Cartesian label is from the left, i.e., a*b*c = (a*b)*c, that is, a*b is expanded first.

4.1.10 character mode bit string: A sequence of alphanumeric characters, i. e. "0" or "1", that represents a string of binary digits. See also **bit field**.

4.1.11 compound data field: A field comprising one or more subfields each containing an elementary data element.

4.1.12 corresponding binary tree: A binary tree which represents the structure of an ordered, rooted tree. The first (leftmost) offspring of a node in an ordered rooted tree becomes the root of the left subtree of the binary tree and the siblings of the node are connected through a series of right-hand links. The preorder traversal sequence of an ordered, rooted tree and its corresponding binary tree are identical.

4.1.13 data descriptive field: A tagged field, residing in the DDR, containing the ISO/IEC 8211 data description for all data fields in the file associated with the same tag.

4.1.14 data descriptive file; DDF: A file containing a data descriptive record and its companion data records.

4.1.15 data descriptive record; DDR: A logical record that contains the control parameters and data definitions necessary to interpret its companion data records. The data descriptive record is the first logical record of a file.

4.1.16 data field: A tagged field which contains user data and is located in a DR.

4.1.17 data record; DR: A logical record containing user data.

4.1.18 delimited structure: A data structure composed of a collection of data elements that are separated by delimiters.

4.1.19 delimiter: A single character that separates data elements and data subfields. See 6.4.1, Table 1 for the use of delimiters.

4.1.20 directory: A field containing an array of identifiers and references to corresponding items of data. It is terminated by a field terminator.

4.1.21 directory entry: A fixed-length field within the directory that contains information about the field tag, the length and the location of a specific field within a given record.

4.1.22 elementary: Having the property of being indivisible without loss of meaning.

4.1.23 entry map: A field in the leader that is used to indicate the structure of the entries in the directory.

4.1.24 escape character; ESC: A control character which is used for code extension purposes. Ii causes the meaning of a limited number of bit combinations following it to be changed. The use of ESC is specified in ISO 2022.

4.1.25 external file title: A string of characters that provides a displayable descriptive title for the interchange file. This need not be the same as the host system file name.

4.1.26 field: A generic term used to describe a contiguous set of octets containing one or more values. The components of a field are referred to generically as subfields. See also **subfield** 2).

4.1.27 field tag; tag: A character string in a directory entry used to identify a data field or an associated data descriptive field.

4.1.28 field terminator; FT: A character, (1/14), used to terminate a variable-length field within a record.

4.1.29 file: A collection of related records treated as a unit.

4.1.30 forest: An ordered set of disjoint, ordered, rooted trees.

4.1.31 interchange format: A format for the exchange, as opposed to the local processing, of data.

4.1.32 interchange level; level: The designation of a prescribed subset of the requirements of this International Standard.

4.1.33 leader: A fixed-length field that occurs at the beginning of each record and provides parameters for the processing of that record.

4.1.34 least significant octet first; LSOF: An ordering of octets in which the least significant octet is placed closest to the beginning of a file.

4.1.35 location; position: The octet count to the position of the first octet of a field. Locations in the leader and directory are relative to the first (0) octet of the leader, and the locations of fields are relative to the base address of data.

4.1.36 map; to map: To establish the correspondence between the elements of two different data structures.

4.1.37 media record: The physical record which is characteristic of a medium and which is written on the media by a single system level write statement, e.g., a block or a sector or a transmission packet. (see also **record**)

4.1.38 most significant octet first; MSOF: An ordering of octets in which the most significant octet is placed closest to the beginning of a file.

4.1.39 n-tuple: A vector, each instance of which has n ordered elements.

4.1.40 null: The non-occurrence of an entity, usually a data element, string or set.

4.1.41 numeric array descriptor: A sequence of numbers which specifies the number of dimensions and extent of each dimension of an array.

4.1.42 ordered, rooted tree; tree; hierarchy: A data structure comprising a set of nodes having directed links such that one node, the root node, has no incoming links, the remaining nodes have precisely one incoming link, and the order of multiple outgoing links from a node is significant. The nodes are referred to as parent nodes and offspring nodes and link direction is from parent node to offspring node.

4.1.43 position: see **location**.

4.1.44 preorder traversal sequence: A sequence of the nodes of an ordered, rooted tree which is produced by the following recursive algorithm:

a) Enter the tree at the root node.
b) Traverse the left-most subtree not previously traversed,
   if no such subtree exists, stop.
c) If b) is not possible, return to the node superior to the subtree and go to b).

4.1.45 record; logical record: A contiguous string of octets in an interchange file, the first five being the record length (see also **media record**).

4.1.46 record length: A data element the value of which is equal to the length in octets of the record.

4.1.47 recursive tree: A tree in which a node appears in its own subtree.

4.1.48 relative position; RP: The position of an octet expressed as a decimal integer relative to the beginning of a record, field or subfield. The relative position of the first octet is numbered "0."

4.1.49 signed binary integer: A two's complement binary number.

4.1.50 subfield: 1) A contiguous string of octets in a field whose position, length and data type are described in the field data description; 2) generically, a component of a field.

4.1.51 subfield label: A character string used in the data description to identify a subfield and its contents.

4.1.52 tagged field: A contiguous string of octets in a record which is identified by a field tag and whose length and relative position are specified in a directory entry. A tagged field may occur in the DDR or in the DRs.

4.1.53 tag: see field tag.

4.1.54 unit terminator; UT: A character, (1/15), used to delimit several types of subfields within variable-length fields in both DDR and DR.

4.1.55 variable-length field: A field, the length of which may vary from occurrence to occurrence.

4.1.56 vector: A one dimensional data structure each instance of which may have a variable number of elements. (see also **n-tuple}**

4.1.57 vector label: A vector the elements of which are labels used to identify each element in a vector of data elements. A vector label is a special case of a Cartesian label.

4.2 Abbreviations

For the purposes of this International Standard the following abbreviations apply:

BCS	basic character set
DDF	data descriptive file
DDR	data descriptive record
DFD	data field description, see B.2
DR	data record
ESC	escape character
FT	field terminator
FTAM	file transfer and access management
LR	logical record
LSOF	least significant octet first
MSBF	most significant octet first
RP	relative position
SPACE	SPACE character
TAB	horizontal tabulation character
UT	unit terminator
level	interchange level
tag	field tag
0...n	a tag with zero or more leading ZEROs