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

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.

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