International Standard 8211

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

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

Traitement de l'information — Spécification d'un fichier de données descriptif pour l'échange de l'information

First edition — 1985-12-15

UDC 681.3.04 **Ref. No. ISO 8211-1985 (E)**

Descriptors : data processing, information interchange, files, specifications.

Price based on 27 pages

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.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 8211 was prepared by Technical Committee ISO/TC 97, *Information processing systems.*

Users should note that all International Standards undergo revision from time to time and that any reference made herein to any other International Standard implies its latest edition, unless otherwise stated.

Printed in Switzerland

Contents

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

0 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 make. 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, if possible, 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 only and is not intended to be used for general processing.

The aim when developing this International Standard was to define an interchange format into which the sender's information is mapped and is conveyed to the receiver's system. Upon receipt of the information in the interchange format it is then mapped into the receiver's format without loss of structure and content. This International Standard specifies a method for describing a robust interchange structure which can accept most user data structures. The method enables the sender to preserve structure information and to convey it to the receiver with the data so that the receiver can remap the structure and data into the local system.

Most data structures in common use can be described and interchanged using this International Standard. The structures within the interchange file can be of the following forms : elementary data, vectors, arrays and hierarchies. User file structures such as sequential, hierarchical, relational and indices can be converted into the interchange structure. Network structures can be interchanged but additional pre-processing and post-processing is necessary to preserve logical linkages.

This International Standard is medium-independent and requires an environment in which International Standard labels and file structures can be written on or read from the standard media chosen. It is assumed that variable-length records can be processed by the supporting label and file processing system. It requires a computer process capability to map the user file or data base management system 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 this International Standard. The interchange standard requires the use of the ISO 646 coded character set in control fields and permits the use of extended 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 level one and processes multiple fields containing structured user data comprising a variety of data types. The third level supports level two and hierarchical data structures.

NOTE — Additional information concerning the application of this International Standard is given in annex A.

1 Scope and field of application

This International Standard specifies an interchange format to facilitate the transfer of files containing data records between computer systems. It is not designed as a record format for indigenous files of any specific system. It defines a generalized structure which can be used to transmit, between systems, records containing a wide variety of data types and structures. It provides the means for the description of the contents of data records but does not define their contents.

This International Standard specifies

a) media-independent file and data record descriptions for information interchange. It assumes the use of other International Standards for labelling and file structure such as ISO 1001, ISO 4341, ISO 7665;

b) the description of data elements, vectors, arrays and hierarchies containing character strings, bit strings and numeric forms. The numeric forms are specified by ISO 6093;

c) a data descriptive file comprising a data descriptive record and companion data records that enable information interchange to occur with minimal specific external description;

d) the data descriptive record that describes the characteristics of each data field within the companion data records;

e) three levels of interchange depending on the complexity of the allowed structure (see 5.2.1.2).

2 Conformance

Interchange files shall be in conformance with 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 interchange level to which the contents of files conform.

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

3 References

ISO 646, *Information processing — ISO 7-bit coded character set for information interchange.*

ISO 1001, *Information processing — Magnetic tape labelling and file structure for information interchange.*

ISO 2022, *Information processing — ISO 7-bit and 8-bit coded character sets — Code extension techniques.*

ISO 4341, *Information processing — Magnetic tape cassette and cartridge labelling and file structure for information interchange.*

ISO 6093, *Information processing — Representation of numerical values in character strings for information interchange.* [1)]

ISO 7665, *Information processing — File structure and labelling of flexible disk cartridges for information interchange.*

The following document is also relevant to this International Standard :

ISO international register of character sets to be used with escape sequences.

4 Definitions

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

4.1 alphanumeric character : A character occurring in columns 2 to 7 inclusive [except position (7/15)] of the International Reference Version of ISO 646.

NOTE — The characters specified in this International Standard are represented by their position (column/row) in the coded character set table of ISO 646, or by their acronyms, i.e., ESC, RS, US, and SP.

4.2 array descriptor : A sequence of numbers which specifies the number of dimensions and extent of an array.

4.3 base address of data : A data element the value of which is equal to the byte count up to the first data field following the field terminator of the directory, where the specified origin (0) is the first byte of the leader.

4.4 bit field : A data field comprising only binary digits and, when necessary, filled on the right with binary zeros to complete a byte. See also character mode bit string.

4.5 byte : A collection of n bits.

NOTE — This International Standard is media-independent and the number of bits will be media-dependent.

4.6 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 $\vec{a}$ and $\vec{b}$ are the vector labels, $\vec{a} = [a(1), \ldots, a(n)]$ and $\vec{b} = [b(1), \ldots, b(m)]$, then the Cartesian label, $\vec{a} \cdot \vec{b} = [a(1)b(1), a(1)b(2), \ldots, a(1)b(m), \ldots, a(n)b(m)]$ where $a(i)b(j)$ is a concatenation of $a(i)$ and $b(j)$ which forms an identifier of the i,j element of a corresponding data array.

4.7 character mode bit string : A sequence of characters (0 or 1) that represents a string of binary digits. (See also bit field.)

4.8 compound data field : A field comprising one or more elementary data elements.

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

4.10 data descriptive record; DDR : A record that logically precedes the data records and contains the control parameters and data definitions necessary to interpret companion data records. The data descriptive record is the first logical record of a file other than the file labels (if applicable).

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

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

4.13 delimiter : A single character that separates data elements and data fields. (See table 1 for the use of delimiters.)

1) At present at the stage of draft.

4.14 directory : A table of identifiers and references to corresponding items of data.

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

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

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

4.18 escape character; ESC : A control character which is used to provide additional characters. It alters the meaning of a limited number of contiguously following bit combinations.

The use of this character is specified in ISO 2022.

4.19 field terminator; FT : A character used to terminate a field within a record, (1/14) in ISO 646.

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

4.21 file title : A string of characters that provides a displayable descriptive title for the interchange file.

This need not be the same as the file name.

4.22 hierarchy; hierarchical structure : A rooted, ordered tree structure comprising a superior root node with successive multiple ordered subtrees at increasingly inferior nodes, ultimately terminating in leaf nodes.

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

4.24 interchange format : A format for the exchange, as opposed to the local processing, of records.

4.25 label : A character string used to identify or name a field or subfield and its contents.

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

4.27 location : The byte count to the position of the first byte of a field.

Locations in the leader and directory are relative to the first (0) byte of the leader, and the location for data descriptive fields and user data fields are relative to the base address of data.

4.28 logical record; record : A collection of related data elements independent of their representation on a medium.

4.29 to map : To establish the correspondence between the elements of two structures.

4.30 null : Pertaining to the condition of non-occurrence of an entity, usually a data element, string or set.

4.31 preorder traversal sequence : A sequence of the nodes of a hierarchy produced by the following recursive algorithm :

a) enter the tree at the root node;

b) traverse the left-most subtree not previously traversed;

c) if b) is not possible, return to the node superior to the subtree and go to b).

4.32 record length : A data element the value of which is equal to the length in bytes of the record.

4.33 relative position; RP : The position of a byte expressed as a decimal integer relative to the beginning of a field.

The first relative position is numbered "0".

4.34 tag : An identifier in a directory entry used to specify the internal name of an associated field.

4.35 unit terminator; UT : A character used to delimit several types of subfields within variable-length fields in both DDR and DR, (1/15) in ISO 646.

4.36 variable-length field : A field the length of which varies from occurrence to occurrence.

4.37 vector label : A vector the elements of which are labels (i.e., "column" headings or "row" headings) used to identify each element in a vector of data elements.

5 Interchange file

5.1 General structure

This subclause specifies the general structure of the interchange file and subsequent subclauses provide the detailed specifications. Figure 1 shows a schematic representation of a file and the file labels.

ISO standard file labels	
Data Descriptive File :	Data Descriptive Record Data Records
ISO file termination indicator	

Figure 1 — File and file label schematic representation

This International Standard specifies multiple Data Descriptive Files (DDFs), each comprising logical records with the required ISO interchange file labels or headers for the particular medium. Each file shall consist of the following logical records :

a) the Data Descriptive Record (DDR), and

b) the Data Records (DRs).

The overall structure shall be as shown in figure 2, which is an expanded logical schematic representation of the DDR and DR, the leaders and directories of each record, typical records and a typical data field for each record. The DDR and DR records shall have the same leader, directory, field and record structure although their contents vary. A provision is made to omit the repetition of identical DR leaders and directories for the interchange for repetitive, fixed-format data.

NOTES

1 The logical juxtaposition of fields is shown and the meaning of some pointers and field lengths is indicated. For a physically sequential medium, figure 2 represents its physical order.

2 The special field tags specified in this International Standard are described in the following format : 0...n where "n" is a decimal number and "0..." implies sufficient zeros on the left to fill the tag field.

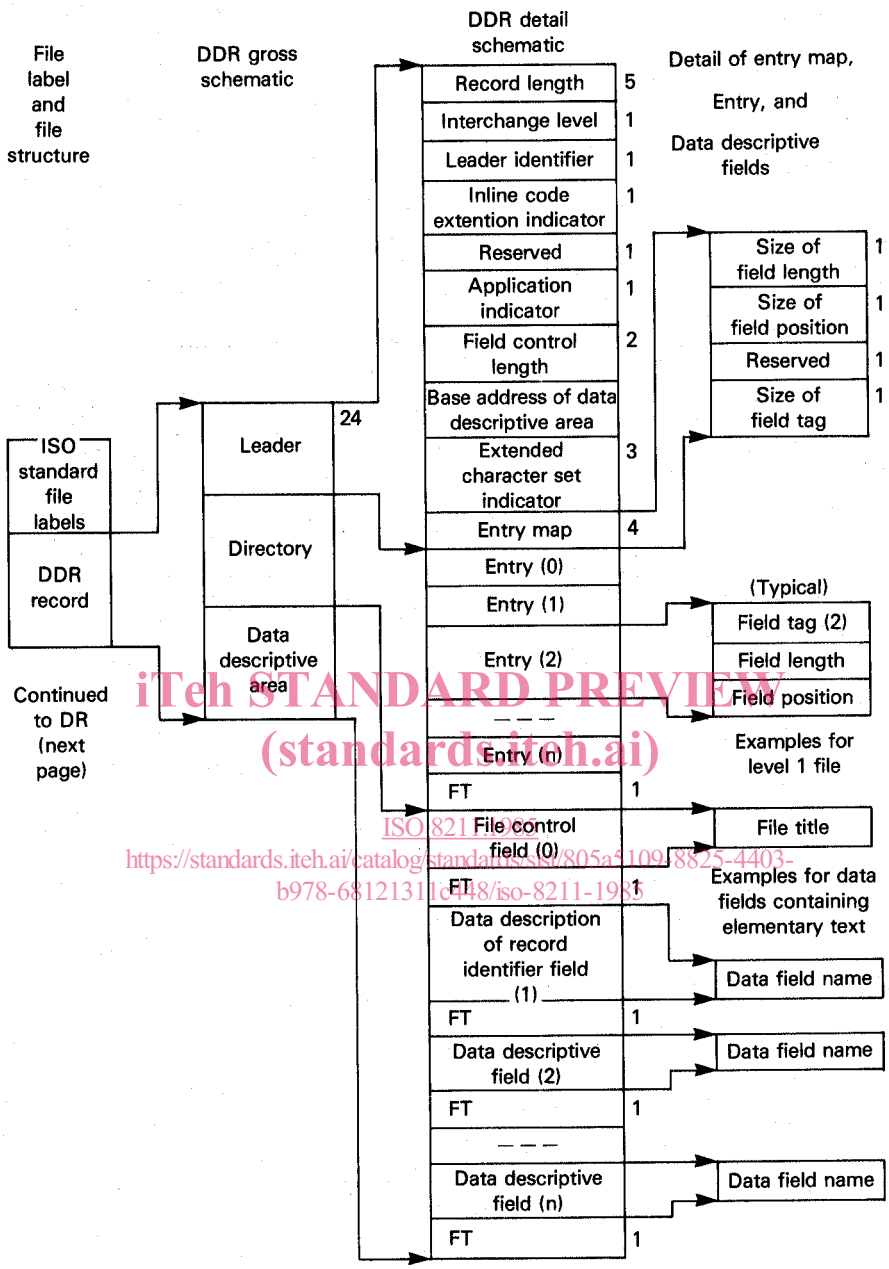
Each data descriptive field of the DDR contains a data description of the user data in a DR user data field having the same tag. The DDR (but not the DR) has a special tag 0...0 and a corresponding field which contains field controls, an optional file title and in the case of a hierarchy, structure information. The DRs have a special field to identify a record and the DDR contains a description of this field in the data descriptive field having the same tag (0...1). The contents of a DDR variable-length field vary depending upon the values of parameters in the DDR leader.

The DDR data descriptive fields shown in figure 2 are for elementary character data fields and those shown in figure 12 are for compound data fields with all optional subfields included.

NOTES

1 The contents of a DR user data field may be highly varied, depending upon its description in the DDR; no example is given in figures 2 and 12. See annex B for examples of data fields.

2 Throughout the remainder of this International Standard the length of fields, except for bit fields, is given in bytes whose length in bits may be media-dependent. The contents of fields are referred to as characters, and multiple byte character sets are permitted in user fields (see 7.1.4). Therefore in these cases the field length is not equal to the number of characters.



See figure 12 for examples of data descriptive fields for level 2 files containing compound data fields.

NOTES

1 Separator characters
FT = Field Terminator

2 Fixed field lengths are given in bytes to the right of each field.

3 Tags, their corresponding entries and fields are designated by (i).

Figure 2 — Expanded logical schematic representation of data descriptive file

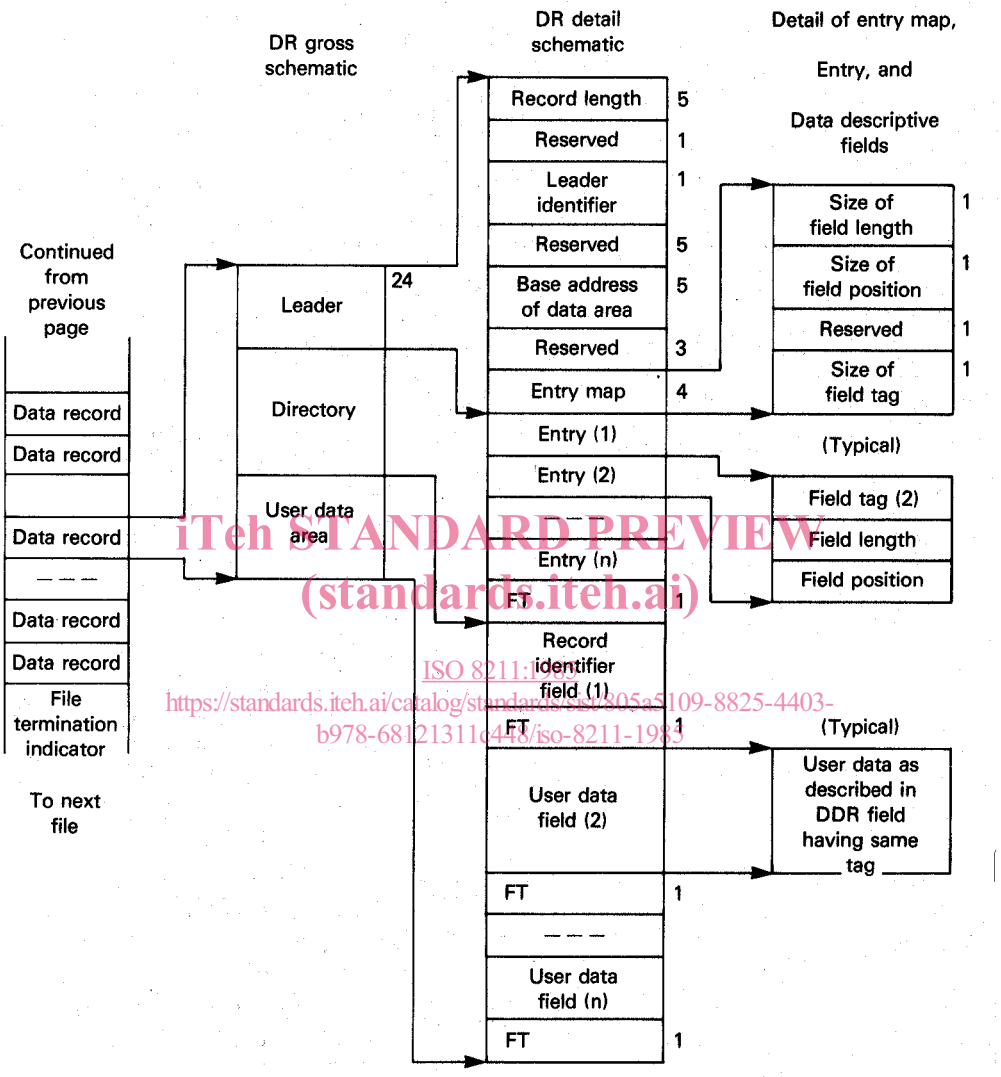


Figure 2 — Expanded logical schematic representation of data descriptive file (*concluded*)

5.2 Data Descriptive Record (DDR)

The DDR shall consist of the areas and terminators shown in figure 3 and shall be the first record of the file.

Name of area	Length
Leader	24
Directory	$k \times p$
Field terminator	1
Data descriptive area	variable
Field terminator	1

Figure 3 — DDR schematic

Each of the logical records shall consist of

a) a leader of 24 characters;

b) a directory of length $k \times p$, terminated by a field terminator, (1/14), where k is the number of directory entries and p is the length of each entry (see 5.2.2); and

c) a set of k variable-length fields, each terminated by a field terminator, (1/14).

5.2.1 DDR leader

The DDR leader shall consist of the fields shown in figure 4 and is further specified in 5.2.1.1 to 5.2.1.9.

RP	Name of field	Length	Contents
0	Record length	5	digits
5	Interchange level	1	digit
6	Leader identifier	1	character
7	Inline code extension indicator	1	character
8	Reserved[1)]	1	SPACE character
9	Application indicator	1	character
10	Field control length	2	digits
12	Base address of data	5	digits
17	Code character set indicator	3	characters
20	Entry map	4	digits

1) *Reserved for future standardization.*

Figure 4 — DDR leader schematic

5.2.1.1 Record length field (DDR RP 0 to 4)

This field shall specify the total length of the DDR in bytes. The contents of this field shall be digits. A DDR length of 0 will signify a length in excess of 99 999.

5.2.1.2 Interchange level field (DDR RP 5)

This field shall specify the level of the interchange file as specified in clause 8. The content of this field shall be the digit 1, 2 or 3.

The value

1 shall mean that the file conforms to a level 1 file;

2 shall mean that the file conforms to a level 2 file;

3 shall mean that the file conforms to a level 3 file.

A level 1 file shall contain elementary character data fields (see 6.1) but not compound data fields nor hierarchical structures. A level 2 file shall contain compound data fields (see 6.2) but not hierarchical structures. A level 3 file shall contain compound data fields and a list of the tag pairs (see 5.2.3.1.2) describing the hierarchical structures.

5.2.1.3 Leader identifier field (DDR RP 6)

This field shall specify that this record is the DDR and shall contain the character "L".

5.2.1.4 Inline code extension indicator (DDR RP 7)

This field shall indicate if inline escape sequences are used in data fields to designate extended coded character sets as specified in ISO 2022.

The value

SP shall mean that no extensions are used;

E shall mean that extensions are used.

5.2.1.5 *Reserved for future standardization* (DDR RP 8)

This field is reserved for future standardization.

5.2.1.6 Application indicator field (DDR RP 9)

This field is reserved for future standardization and shall contain the character SPACE.

5.2.1.7 Field control length field (DDR RP 10 and 11)

This field shall specify the number of bytes of the data descriptive field devoted to data element type and structure codes, delimiters, and other positions reserved for future standardization (see 5.2.3.1.1).

The contents of this field shall be the digits 00, 03, 06 or 09 (see 5.3.3.2 and 7.1.3).

5.2.1.8 Base address of data descriptive area (DDR RP 12 to 16)

This field shall specify the position of the first data descriptive field of a DDR.

NOTE — The first data descriptive field will be the file control field or the record identifier field.

The contents of this field shall be digits and shall be equal to the combined length in bytes of the leader and directory including the field terminator at the end of the directory.

5.2.1.9 Extended character set indicator field (DDR RP 17 to 19)

This field shall specify the use of default coded character set extensions in the file.

The values in this field shall have the following meanings :

a) (2/0)(2/0)(2/0) : Only the International Reference Version of ISO 646 character set has been designated as the default for the file.

b) (2/0)(2/1)(2/0) : Extended character sets have been designated as the default for one or more data fields (see 7.1.1).

c) Truncated escape sequence : An extended character set has been designated as the default for the entire file (see 7.1.2).

5.2.1.10 Entry map field (DDR RP 20 to 23)

This field specifies the lengths of directory entry subfields and shall consist of the subfields shown in figure 5 (see 5.2.2).

Each subfield of this field shall contain a single digit.

RP	Name of field	Length	Contents
20	Size of field length	1	digit
21	Size of field position	1	digit
22	Reserved for future standardization	1	digit
23	Size of field tag	1	digit

Figure 5 — DDR entry map schematic

5.2.1.10.1 Size of field length field (DDR RP 20)

This field shall specify the size in bytes of the field length subfield of the directory entries and shall be a digit in the range of "1" to "9" inclusive.

5.2.1.10.2 Size of field position field (DDR RP 21)

This field shall specify the size in bytes of the field position subfield of the directory entries and shall be a digit in the range of "1" to "9" inclusive.

5.2.1.10.3 Reserved for future standardization (DDR RP 22)

This field is reserved for future standardization as an extended entry map and shall be the digit "0".

5.2.1.10.4 Size of field tag field (DDR RP 23)

This field shall specify the size in bytes of the field tag subfield of the directory entries and shall be a digit in the range of "1" to "7" inclusive.

NOTE — For the purposes of 5.2 the following nomenclature is used :

m is the size of the field length subfield;

n is the size of the field position subfield;

t is the size of the field tag subfield.

5.2.2 DDR directory

The DDR directory shall consist of repeated DDR directory entries, the subfield lengths of which shall be specified in the entry map. The DDR directory shall contain one DDR directory entry for each data descriptive field and shall end with a field terminator, (1/14). The DDR shall define all DR tags.

The DDR directory entry specifies the location and length of a corresponding data descriptive field and shall consist of the subfields shown in figure 6. Each entry shall contain a field tag, field length and field position, in that sequence and shall consist of $m + n + t$ bytes.

The DDR directory entries shall be in one to one correspondence with the data descriptive fields. For hierarchical data structures (see 5.2.3.1.3), the directory entries of the DDR shall be in the same order as the pre-order traversal sequence of the generic data tree.

RP	Name of field	Length	Contents
$p(i-1)$	Field tag	t	alphanumeric
$p(i-1)+t$	Field length	m	digits
$p(i-1)+t+m$	Field position	n	digits

where

$p = t + m + n$ and;

i = the index of the directory entry.

Figure 6 — DDR directory entry schematic

5.2.2.1 DDR field tag field

This field shall contain a field tag identifying a data descriptive field and shall consist of between one and seven alphanumeric characters. The same field tag shall occur only once within the DDR.

5.2.2.1.1 Field tag 0...0 shall identify the optional file control field and if present shall occur only in the DDR.