

INTERNATIONAL STANDARD

ISO
6156

First edition
1987-02-01



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

Magnetic tape exchange format for terminological/lexicographical records (MATER)

*Format d'échange sur bande magnétique des données terminologiques et/ou
lexicographiques (MATER)*

(standards.iteh.ai)

ISO 6156:1987

<https://standards.iteh.ai/catalog/standards/sist/823b5e47-3f6b-4486-8523-f93c0e490921/iso-6156-1987>

Reference number
ISO 6156:1987 (E)

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 6156 was prepared by Technical Committee ISO/TC 37, *Terminology*.

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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

<https://standards.iteh.ai/catalog/standards/sist/62565c47-3f6b-4486-8523-f93c0e490921/iso-6156-1987>

Magnetic tape exchange format for terminological/lexicographical records (MATER)

1 Scope

This International Standard specifies the requirements for a generalized format, the definition and the layout of terminological/lexicographical data on magnetic tape as well as the meaning and the layout of the tags associated with each element.

It describes an interchange principle between data processing systems but does not aim at defining how or in which form a user manages his own system.

The annexes are examples which illustrate possible applications of this International Standard.

ISO 1863, *Information processing — 9-track, 12,7 mm (0.5 in) wide magnetic tape for information interchange recorded at 32 rpm (800 rpi)*.

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

ISO 2709, *Documentation — Formats for bibliographic information interchange on magnetic tape*.¹⁾

ISO 3788, *Information processing — 9-track, 12,7 mm (0.5 in) wide magnetic tape for information interchange recorded at 63 rpm (1600 rpi), phase encoded*.

ISO 3166, *Codes for the representation of names of countries*.

2 Field of application

This International Standard is designed for information interchange on magnetic tape. The information concerned consists of various monolingual or multilingual terminological or lexicographical data used for this type of exchange. It is also meant to enable users implementing a set of mutually compatible data categories to interchange terminological/lexicographical information with the aid of data processing techniques.

ISO 5426, *Extension of the latin alphabet coded character set for bibliographic information interchange*.

ISO 5427, *Extension of the Cyrillic alphabet coded character set for bibliographic information interchange*.

ISO 5428, *Greek alphabet coded character set for bibliographic information interchange*.

3 References

ISO 639, *Symbols for languages, countries and authorities*.

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

ISO 962, *Information processing — Implementation of the 7-bit coded character set and its 7-bit and 8-bit extensions on 9-track 12,7 mm (0.5 in) magnetic tape*.

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

ISO/R 1149, *Layout of multilingual classified vocabularies*.

4 Definitions

4.1 terminological/lexicographical record: A collection of data fields, including a record label, reference data, a directory and terminological or lexicographical data of one terminological/lexicographical unit (e.g. entry in a dictionary).

4.2 directory: An index to the location of data fields within a record, containing the tag, length, location and specifier of each data field within the record.

4.3 data field: A portion of variable length of a record containing a particular category of data following a directory and associated with one entry of the directory.

1) The format of ISO 6156 is partly based on ISO 2709 and adapted to terminological and lexicographical purposes.

4.4 data element (for the purpose of this International Standard): The smallest separately identifiable portion of a record used for the description and/or representation of terminological and/or lexicographical data.

4.5 data category: Uniquely defined type of a terminological/lexicographical data element which can be used to structure and describe the content of a terminological/lexicographical record.

4.6 catalogue of data categories: The set of data categories used to describe terminological and/or lexicographical data.

4.7 tag: Three characters associated with a data field and used to identify it.

4.8 terminological/lexicographical unit: A set of data fields containing all the data either on a single term (and its various meanings) or on a single concept (and all the synonymous terms related to it).

NOTE — Figure 1 illustrates the structure of the interchange file.

5 Character set

The character sets used shall be those characters in the international reference version of ISO 646, together with, if necessary, other ISO standard character sets devised in accordance with the provisions of ISO 2022, such as ISO 5426, ISO 5427 and ISO 5428.

6 Interchange file structure

MATER defines the file format for the interchange of terminological/lexicographical data (see clause 7). These are hereafter referred to as interchange data.

If other types of file are to be exchanged, bibliographic data shall be transferred to a separate file and formatted in accordance with the UNISIST Reference Manual of IFLA's UNIMARC specifications (see 6.2.3) and any other data shall be transferred to supplementary auxiliary files (see 6.2.4).

Any other information the user may need for the purpose of processing interchange files but which cannot be transferred on the file itself must be entered in the work sheet (see annex A).

6.1 Files and labels

All files shall be labelled according to ISO 1001. Of the possible levels of labelling given in ISO 1001, level 3 is chosen which applies to variable-length records. Therefore, the following headers shall be used:

VOL 1	Volume Header Label
HDR 1	First Header Label
HDR 2	Second Header Label
EOV/EOF 1	First End of Volume/File Label
EOV/EOF 2	Second End of Volume/File Label

No other labels shall be used within this Standard. The content of the labels is laid down in ISO 1001. All records are of variable length.

6.2 Interchange file

The structuring of the interchange file into interchange units and records (see figure 1) is provided in order to allow the user to build up hierarchically structured files.

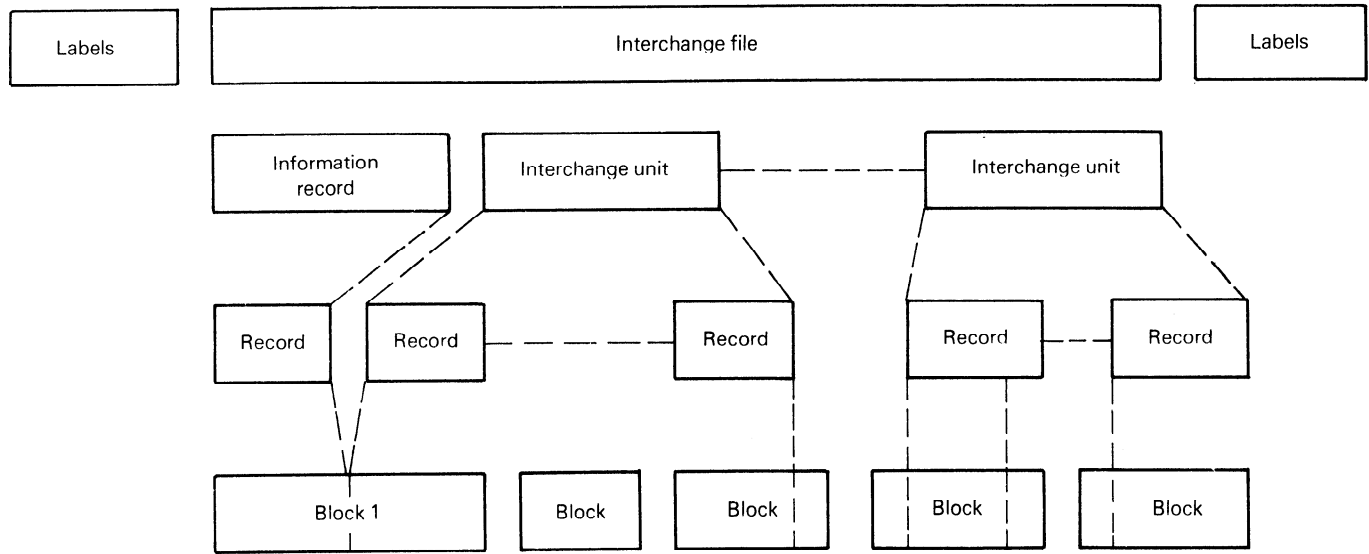


Figure 1

6.2.1 Interchange units

The interchange file consists of a number of interchange units. Each interchange unit can be made up of a number of records of variable length (see 7.1). Any one block should not exceed a length of 2048 characters.

6.2.2 Information record

The information record is intended to enable the receiving agency or institution to proceed at once with the automatic analysis of the interchange file. Apart from general information (e.g. the name of the producer and the creation date) the information record shall also contain the the separating characters used, the name of the character set used (according to the ISO International register of character sets to be used with escape sequences) and a character set reference table. The use of the information record is optional. If it is used it must be the first record in the file. To identify this record uniquely — if it is used — all eight bits of byte 5 (position 2 of the record label = status) should be set to 1.

6.2.3 Bibliographic data file

A separate file should be used for exchanging bibliographic data relating to data transferred. The record structure of the data in this file shall be in accordance with either of the following two implementations of ISO 2709:

a) the UNISIST reference manual for machine-readable bibliographic descriptions;

b) The IFLA universal MARC format.

6.2.4 Auxiliary files

Such files can include additional information relevant to the evaluation of interchange data, e.g. digitalized illustrations, textual material which can be used for processing data, DP-programs, etc.

The record structure is not fixed. It must be agreed between the interchanging institutions or defined in an annex to the work sheet.

7 Interchange unit structures

This clause only defines the logical record structure. In particular this clause does not specify any specific system of record writing. This format is partly based on ISO 2709 and adapted to terminological and lexicographical ends.

An interchange unit contains the data of one terminological/lexicographical record (e.g. entry in a dictionary), and normally comprises a record of variable length in the interchange file. In some cases, the interchange unit may consist of several records. Records belonging to the same interchange unit shall have the same identification number, the individual records being distinguished from each other by sequential numbering in the record count.

Identification number: Positions 0 to 7 (see 7.2.2 "Reference data").

Record count: Positions 8 and 9 (see 7.2.2 "Reference data").

Record counting is included to facilitate the application of this International Standard to small computer systems.

7.1 Record types

An interchange unit based on the MATER format can either be a single record or consist of a number of consecutive records.

The following interchange unit consists of a single record without any overflow (record count = 00):

Reference data	
Positions 0 to 7	00000027
Positions 8 and 9	00

Identification No.

Record count

Figure 2

If an interchange unit spans a number of records, these are divided into two categories under the MATER format:

a) Main records;

b) Overflow records.

The first record is the main record; the remainder are overflow records.

Example: Interchange unit consisting of three records (record count of main record = 01):

1st Record = Main record

Positions 0 to 7	00000028	Identification No.
Positions 8 and 9	01	Record count

2nd Record = Overflow record

Positions 0 to 7	00000028	Identification No.
Positions 8 and 9	02	Record count

3rd Record = Last Overflow record

Positions 0 to 7	00000028	Identification No.
Positions 8 and 9	99	Record count

Figure 3

The last overflow record should be indicated by the record count of 99.

If an interchange unit contains many items of information in a number of languages, it may be expedient, instead of setting corresponding items of data in the various languages side by side, to bring all the information in each language together and only then to link the composite data in all the various languages (see 7.2.1 note 4).

It may be necessary to break down an interchange unit into a number of records for the following reasons:

- a) The total length of the interchange unit and its directory may be greater than 2 044 bytes. In such cases, the unit must be broken down into a main record and overflow records in such way that no data field begins in one record and ends in the next, i.e. the end-of-record must always coincide with an end-of-field. The data fields must be so matched to the records as to make the fullest possible use of the natural breaks in the structure of the interchange unit.
- b) A separate record must be made for every language of the interchange unit.

7.2 Structure of the format

The general structure of a record is shown in figure 4.

Record label
Reference data
Directory
Data fields
Record separator (IS ₃) ¹⁾

1) IS₃ is the record separator of ISO 646.

Figure 4

A record includes the items defined in clause 4 and contains the following fixed and variable length fields:

- a record label : fixed length;
- reference data : fixed length;
- a directory : variable length;
- interchange data fields : variable number and variable length;
- field separator(s) : one character;
- a record separator : one character.

7.2.1 Record label

The record label section constitutes the first section of the record and carries the key information for identifying and processing the record.

The elements contained in a record label are shown in table 1.

Table 1

No.	Character positions	Data elements		Meaning	Content
		Length in characters	Type ¹⁾		
1	0 to 4	5	N	User's record length data	Numbers of bytes in the record
2	5	1	C	Record status	N = new A = amended D = deleted
3	6 to 9	4	C	Implementation codes ²⁾	0000 (or codification to be defined)
4	10	1	N	Indicator length	3
5	11	1		Not used	0
6	12 to 16	5	N	Base address of data ³⁾	
7	17 to 19	3		For future use	000
8	20	1	N	Length of "length of data field" for each entry in the directory	4
9	21	1	N	Length of "starting character position" for each entry in the directory	5
10	22	1	N	Specifier length ⁴⁾	3
11	23	1		For future use	0

1) C = left-justified alphanumerical
N = right-justified decimal numerical characters.

2) May indicate the record type: terminological, lexicographical, etc.

3) The position of the first field in the data field section relative to the first byte of the record.

4) This subfield of the directory entry (the "implementation-defined part" provided for in ISO 2709) is here used as a specifier of the data in the corresponding data field. The third position indicates the grouping of data fields, the first and the second positions indicate the language of the data (see 7.2.4.1). One or the other of these specifiers is set to zero if it is not used.

7.2.2 Reference data

This section consists of a fixed field of a total length of 96 bytes. It is located immediately after the record label. The location of this section was introduced in deviation from ISO 2709 in order to enable more convenient (i.e. immediate) access to the exchange unit as a whole (e.g. to select a subset of data from the complete record or to have direct access to the identification number for selection purposes etc.).

Table 2

No.	Character positions	Data elements		Meaning	Content
		Length in characters	Type ¹⁾		
1	0 to 7	8	N	Identification number of interchange unit	Serial numbers of units in the file are in ascending order commencing at decimal 00000001
2	8 and 9	2	N	Record count (within interchange unit)	See 7.1
3	10 to 17	8	C	Originating agency's internal identification number of the interchange unit	Allocated by the originating agency
4	18 to 23	6	N	Creation date of the interchange unit	Date of last change by the originating agency (YYMMDD)
5	24 to 47	24	C	Subject field code(s)	The meaning of the codes is to be given in the work sheet
6	48 to 71	24	C	Language(s) of the exchange unit	The codes of ISO 639 should be used
7	72 to 95	24		Reserved for future use	

- 1) C = left-justified alphanumerical
N = right-justified decimal numerical characters.

7.2.3 Directory

The directory contains information about the data recorded in the data fields and facilitates the processing operation. The directory covers only the data in the record of which it is part, and not the data in the overflow records, each of which has its own directory. The entries in the directory are all of identical length (15 bytes). However the number of entries will vary from record to record, being identical to the number of data fields of the record concerned.

The structure of a directory entry shall be as show in figure 5:

Positions 0 to 2	Tag
Positions 3 to 6	Field length
Positions 7 to 11	Starting character position of data field
Positions 12 to 14	Specifier (see 7.2.1, note 4)

Figure 5

The sequence of entries in the directory shall be the same as the sequence of the corresponding data fields in the record.

NOTES

- 1 The starting character positions given in the directory are relative to the data base address (specified at character positions 12 to 16 in the record label). The starting character position of the first field is 0.
- 2 The directory always ends with a field separating character (IS₂ of ISO 646).
- 3 The same tag may be repeated as required.
- 4 The field length indicates the total length of the data field, including the indicator and the field separator.

7.2.4 Data fields

The data fields are used for storing the interchange data. Each field contains data of only one category. The field length is variable.

The structure of a data field is shown in figure 6.

Positions 0 to 2	Indicator
Positions 3 to $k-1$	Data
Position k	Field separator (IS ₂)

Figure 6

The tag is repeated, as indicator, in order to facilitate checking of the print-out without recurring to the directory. A data field may not extend beyond the end of record. Where necessary, an overflow record shall be used.

7.2.4.1 Grouping

Several data fields, or groups of data fields, may be repeated with the same tags. The third position of the specifier should be used to differentiate between the repeatable fields, or groups of fields.

7.2.4.2 Examples

<i>Example 1:</i>	Specifier	Tag
	LL 0	100
	LL 1	502
	LL 2	502
	LL 3	502

<i>Example 2:</i>	Specifier	Tag
	LL 0	100
	LL 1	502
	LL 1	201
	LL 1	202
	LL 2	502
	LL 2	201
	LL 2	202
	LL 3	502
	LL 3	202

7.2.5 Separators

The separators defined in ISO 646 shall be used. These separators are the record separators (IS₃) and the field separators (IS₂).

8 Utilization

Data category codes are used as tags in directory entries and as indicators in data fields.

An example of a data category is shown in figure 7 and the detailed structure of each is shown in figure 8.

9 Work sheet

A printed work sheet should be transmitted with the interchange tape to explain the structure and content of the data to the user.

A layout of the work sheet is proposed in annex A, as an example.

10 Interchange data categories

For information interchange purposes, each sending agency should define the data categories in order to structure the interchange units. The annexes can be used as a guide for so doing.

The tags used within an interchange file should be constructed in the following manner:

— the first digit of the tag for each data category shall conform to the numbering of the classes as laid out in table 3;

— the second and third digits should be either the digits given in the annexes if the categories described there can be used or two alphanumerical characters if the sender's data cannot be described by the categories of the annexes. In the latter case each tag used within the interchange file shall be described fully in the work sheet.

A list of classes of data categories is given in table 3.

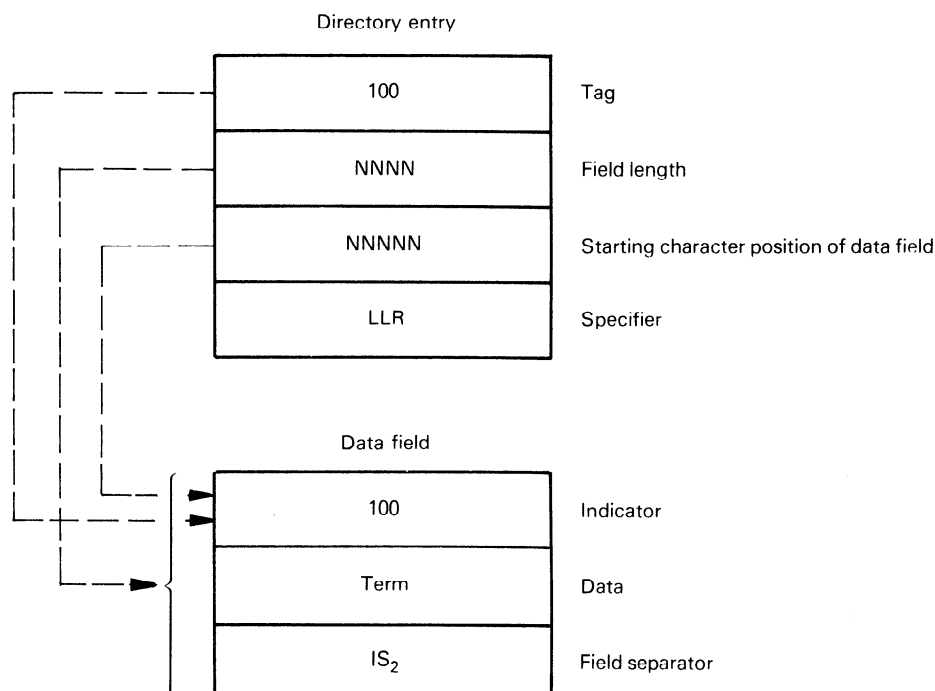


Figure 7

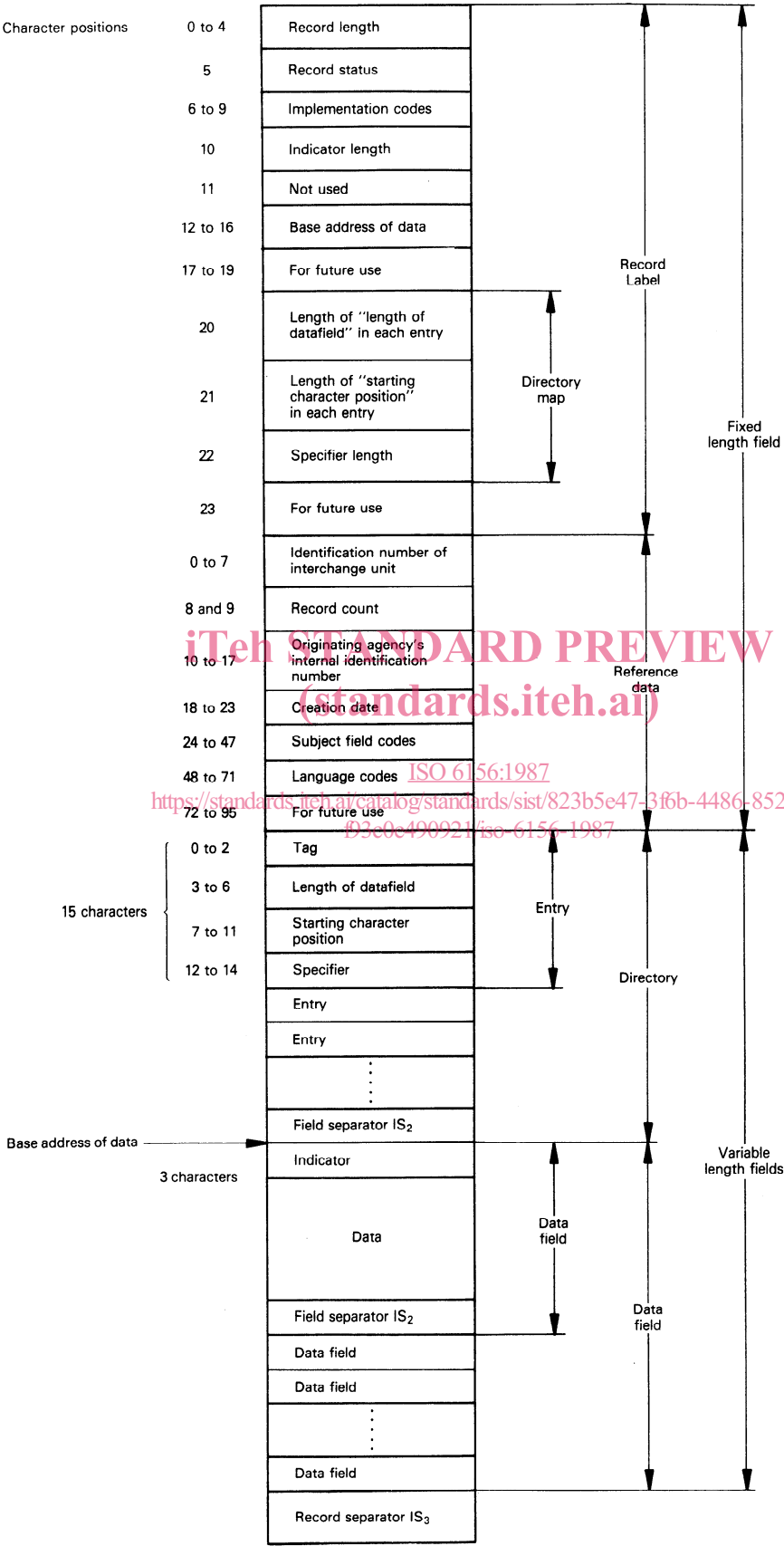


Figure 8

Table 3

Class	Repeatability	Nature of the data	Observations
1..		Terminological/lexicographical entry	
2..	Yes	Morphological description	
3..	Yes	Subject field and description of the entry for the record concerned	
4..	Yes	Definitions, examples, contexts of the entry, for the record concerned	
5..	Yes	Relationships between concepts and/or terms in the same language	Broader terms, narrower terms, homonyms, synonyms, related terms... in the language of the entry
6..	Yes	For future use	
7..	Yes	Other information	
8..	Yes	Sources	Associated with any of the data categories
9..	Yes	Author(s) of the interchange unit	Including copyright

iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO 6156:1987

<https://standards.iteh.ai/catalog/standards/sist/823b5e47-3f6b-4486-8523-f93c0e490921/iso-6156-1987>