STANDARD ST.30

Version 1.0

RECOMMENDATION CONCERNING A STANDARD MAGNETIC TAPE FORMAT FOR THE EXCHANGE IN MACHINE-READABLE FORM OF BIBLIOGRAPHIC DATA, ABSTRACTS AND FULL TEXTS OF PATENT DOCUMENTS

Standard adopted during the 1980's

TABLE OF CONTENTS

- Introduction
- References
- Definitions
- Structure of the exchange format
  - Record label
    - Record length[1] (character positions 0 to 4)
    - Record status (character position 5)
    - Reserved positions (character position 6 to 9)
    - Indicator length (character position 10)
    - Identifier length (character position 11)
    - Base address of data (character positions 12 to 16)
    - Trailer records and numbering of trailer records (character positions 17 & 18)
    - Reserved position (character position 19)
    - Directory map (characters positions 20 to 23)
  - Directory
    - Directory entry
    - Tag
    - Length of datafield
    - Starting character position
    - Implementation-defined part
  - Datafields
    - Record identifier field[5]
    - Reserved fields[5]
    - Patent datafields
- **ANNEX**: Definitions of tags used and description of the datafields identified by these tags
  - Introduction
  - Linked Information
    - Survey of standardized tags

Introduction

1. This recommendation specifies the requirements for a generalized magnetic tape format for the exchange of records, containing bibliographic data, abstracts or full texts of patent documents. It does not define the content of individual records but attempts to standardize the assignment of tags, indicators and identifiers and their use in datafields occurring in such records.

2. This recommendation describes a generalized structure designed specially for communications between data processing systems and not for use as a processing format within systems. Although this recommendation is designed for magnetic tape, its structure may also be used for other data carriers.

3. Wherever possible, this recommendation has been based on existing international standards, in particular on existing WIPO Standards and on those elaborated by the International Organization for Standardization (ISO). Due reference has been made to such standards in those parts of the present recommendation for which they are of particular relevance. Standards being of fundamental importance to this recommendation as a whole are given in the following references.

References

4. The following standards are of fundamental importance to this recommendation:

   (a) WIPO Standard ST.3 – Two–Letter Code for Countries, Organizations and the Like;

   (b) WIPO Standard ST.9 – Recommendation concerning Bibliographic Data on and relating to Patent Documents;

(c) WIPO Standard ST.8 – Standard Recording of International Patent Classification Symbols on Machine–Readable Records;

(d) ISO 1001, Information processing – Magnetic tape labelling and file structure for information interchange;

This document relates to information interchange utilizing magnetic tape, by providing magnetically recorded labels to identify and structure files, and by providing a standard structure for the blocks containing the records that constitute a file. It also specifies a block spanning technique;

(e) ISO 2709, Documentation – Format for bibliographic information interchange on magnetic tape;

This document specifies the requirements for a generalized exchange format which will hold records describing all forms of material capable of bibliographic description as well as related records such as authority records, but does not, however, assign any meaning to tags, indicators or identifiers. The present recommendation closely follows the generalized structure described in ISO 2709 and the terminology used therein;

(f) ISO 8601 – Data elements and interchange formats—information interchange—representation of dates and times;

(g) ISO/R 639-1967 – Symbols for languages, countries and authorities. (This Standard will be applied only insofar as languages are to be coded, as to countries and authorities, see (a), above.)

In addition to the above standards, various other ISO standards are also relevant which deal with the character sets to be used for the recording of letters, figures, punctuation marks, chemical, mathematical and Greek symbols, signs dealing with paragraphs, tables and embedded images, and other special symbols likely to be met with in patent documents.

Definitions

5. For the purpose of this recommendation, the following specific definitions apply:

(a) RECORD: A collection of fields, including a record label, a directory and data referring to one and the same patent document and being treated as one entity.

(b) SUBRECORD: A group of fields within a record which may be treated as an entity.

(c) STRUCTURE: An arrangement of the parts constituting a bibliographic record.

(d) DATAFIELD: A variable length portion of the record containing a particular category of data, following the directory and associated with one entry of the directory.

Note: A datafield may contain one or more subfields.

(e) (field) TAG: Three characters associated with a field and used to identify it.

(f) SUBFIELD: A part of a field containing a defined unit of information.

(g) RECORD LABEL: A field occurring at the beginning of each bibliographic record providing parameters for the processing of the record.

(h) DIRECTORY: An index to the location of the datafields within a record (See paragraphs 19 to 24, below.)

(i) DIRECTORY MAP: A set of parameters specifying the structure of the entries in the directory.

(j) INDICATOR: The first data element, if present, associated with a datafield supplying further information about the contents of the field, about the relationship between the field and other fields in the record, or about the action required in certain data manipulation processes. (See footnote 4 to paragraph 27, below.)

(k) (subfield) IDENTIFIER: A data element, one or more characters in length, immediately preceding and identifying a subfield. (See footnote 4 to paragraph 28, below.)

(l) FIELD SEPARATOR: A control character used to separate and qualify units of data logically, and in some cases hierarchically.

Structure of the exchange format

6. The general structure of a record is shown schematically in Figure 1. A more detailed structure is shown schematically in Figure 2.

Figure 1 – General structure

Record label

Directory
Datafields
Record Separator

7. A record includes the items defined in paragraph 5 and contains the following fixed and variable length fields in the sequence shown in Figure 2:
   - a record label = fixed length
   - a directory = variable length
   - record identifier = variable length
   - reserved fields = variable length
   - patent datafields = variable length
   - field separator(s) = character IS_2 of ISO 646
   - a record separator = character IS_3 of ISO 646.

8. The directory, record identifier, reserved fields and patent datafields are each terminated by a field separator. The record is terminated by the record separator.

Record label

9. The record label shown in Figure 2 is fixed in length and defined as follows:

Record length[1] (character positions 0 to 4)

10. The number of character positions in the record including the record label and record separator. The length is a 5-digit decimal number right adjusted with zero fill if necessary.

Record status (character position 5)

11. A single character describing the status of a record, i.e., new, amended or deleted record. It should be defined by the originator of the record.

Reserved positions (character position 6 to 9)

12. Character positions 6 to 9 are reserved for the internal use of the sender or the recipient of the tape.

Indicator length (character position 10)

13. One decimal digit giving the number of character positions of the indicators.

Identifier length (character position 11)

14. One decimal digit giving the number of character positions of the identifier. The first or only character of this identifier shall always be IS_1 of ISO 646.

Base address of data (character positions 12 to 16)

15. Five decimal digits justified right with zero fill if necessary, and equal to the combined length in characters of the record label and the directory including the field separator at the end of the directory.

Figure 2 – Detailed record structure

Character positions	0 to 4	Record length		Record Label	Fixed length field (24 char)
	5	Record status			
	6 to 9	Reserved positions			
	10	Indicator length			
	11	Identifier length			
	12 to 16	Base address of data			
	17 & 18	Trailer records and numbering of trailer record			

	19	Reserved position			
	20	Length of “length of datafield” in each entry	Directory Map		
	21	Length of “starting character position” in each entry			
	22	Length of “implementation-defined part” in each entry			
	23	For future use			
3 characters		TAG	Entry	Directory	Variable length fields
		Length of datafield			
		Starting character position			
		Implementation defined part (optional)			
		Entry			
		Entry			
Base address of data		Field separator			
Tag 001		Reference data	Record identifier	Data Fields	
		Field separator			
Tags 002 to 009 and 00A to 00Z		Reference data	Reserved fields		
		Field separator			
Additional tags		Indicator	Patent data fields		
		Identifier			
		Data			
		Identifier			
		Data			
		Identifier			
		Data			
		Field separator			
		Indicator			
		Identifier			
		Data			
		Identifier			
		Data			
		Field separator			
		Indicator			
		Identifier			
		Data			
		Data			
		Field separator			
		Record separator			
Next Record					

Trailer records and numbering of trailer records (character positions 17 & 18)

16. Character positions 17 and 18 are used to indicate trailer records and the numbering of trailer records. The total number of trailer records in a set is indicated in character position 18, whilst character position 17 identifies successive trailer records in

the set.

Reserved position (character position 19)

17. Character position 19 is reserved for the internal use of the sender or recipient of the tape.

Directory map (characters positions 20 to 23)

18. These character positions are used as follows:

    (a) Character position 20: One decimal digit equal to the length in characters of the “length of datafield” part of each entry in the directory;

    (b) Character position 21: One decimal digit equal to the length in characters of the “starting character position” part of each entry in the directory;

    (c) Character position 22: One decimal digit equal to the length in characters of the “implementation-defined part” of each entry in the directory;

    (d) Character position 23: Reserved for future use.

Directory

19. The directory consists of a variable number of entries each corresponding to its respective datafield (record identifier, reserved and patent datafields). The directory ends with a field separator (fs).

Directory entry

20. An entry consists of a “tag”, a “length of datafield”, “starting character position”, and “implementation-defined part”, in that sequence.

21. The length of the “tag” is three characters. All entries in a directory have the same structure.

Tag

22. A tag consists of three characters, which, according to the definitions given in the Annex to this recommendation, specify the name of any associated datafield.

Length of datafield

23. This length is either:

    (a) the total number of characters (including indicator(s) and field separator) in the datafield indicated by the preceding tag;

    (b) zero, implying that the directory entry refers to a datafield whose total length is greater than the largest decimal number (n) which can be stored in the “length” of a directory entry. In this case, the datafield is regarded as being divided into a number of parts of which all but the last are of equal length (n). Each part has a corresponding directory entry containing the tag for the datafield and the starting character position of the part to which the directory entry refers. A length “zero” indicates that the directory entry refers to a part of the datafield which is not the final part and that the length of this part is to be taken as (n);

    (c) the number of characters (including field separator) in the final part of a datafield which has been treated as described in (b).

    In the cases described in (b) and (c), all directory entries which refer to parts of the same datafield shall be adjacent and in sequence.

Starting character position

24. A decimal number giving the position of the first character of the datafield identified by the preceding tag, relative to the base address of data [i.e., the starting character position of the first datafield following the directory is 0 (zero)].

Implementation-defined part

25. [To be specified later.]

Datafields

26. All datafields shall end with a field separator. There are three types of fields:

- record identifier field: tag 001[2];
- reserved fields: tags 002 to 009 and 00A to 00Z[3] as required;
- patent datafields: tags 010 to 999 and 0AA to ZZZ as required[4].

Record identifier field[5]

27. Characters identifying the record and assigned by the office creating the record.

Reserved fields[5]

28. A reserved datafield supplies data which may be required for the processing of the record.

Patent datafields

29. Each datafield consists of an indicator, identifier(s), data and a field separator. The lengths of the indicator(s) and identifier(s) are determined by the indicator length and identifier length as defined in the record label which shall be used consistently within each field of the record.

---

[1] The record length described here is a logical record length. For practical reasons relating to machine processing of data, the information may have to be divided into blocks; a standardized technique therefor is specified in ISO Standard 1001. It should be noted that the maximum theoretical record length is 99,999 characters. For recording documents exceeding this length, trailer records have to be created and reference is made to paragraph 16 in this connection.

[2] 0 has the meaning of zero in all these cases.

[3] For alphabetic character use either capital or small letters.

[4] Any combination of numeric and alphabetic characters is allowed; numerals only, alphabetic characters only, or a mixture of both. For alphabetic characters use either capital or small letters. When alpha-numeric tags are used they should never start with 00 since only reserved fields start with two zeros.

[5] Record identifier fields and reserved datafields do not contain indicators or identifiers.

ANNEX: Definitions of tags used and description of the datafields identified by these tags

Introduction

30. The TAGs defined in this Annex refer to bibliographic and textual elements which are considered typical for records on patent documents. These TAGs should always be used as defined.

31. For defining bibliographic, textual or other information elements which are different from those dealt with in this annex and which the Office creating the respective record considers to be suitable, all three-digit combinations of numeric and alphanumeric characters can be used which are not defined in this annex and do not start with “00” (double zero). The content of the fields identified by such non-standardized TAGs should, however, be interpreted by other offices only with the consent and under guidance of the office which created the record. In order to facilitate updating of this annex, the use of non-standardized TAGs starting with two numerals should be avoided (see also following paragraph).

32. For reasons of convenience, the standardized TAGs have been chosen in a way that whenever there is an INID code (according to WIPO Standard ST.9) characterizing the given data element, this INID code coincides with the first two characters of the TAG.

33. If there are specific provisions on how the content of the field identified by a given standardized TAG should be recorded a specific reference is made to the relevant WIPO or ISO Standard.

34. For fields for which no such specific provision exists, the corresponding datafield will be considered as being of variable length.

35. It is recalled that only such TAGs should be used in a record for which the length of the corresponding datafield is different from zero, i.e., for which real data exist. There may *not* be two or more identical TAGs in one and the same record, an exception being in the case of trailer fields used when the specified length is not sufficient.

Linked Information

36. There are fields which, if occurring together in the same record, will have subfields whose information content is cross-linked in a well defined way. For instance, in the case of multiple priorities the respective priority dates, priority countries and numbers of priority application may not be combined arbitrarily. Such interlinked fields are identified by “linked tags” which have been marked in the following paragraphs by a numerical index which is the same for all TAGs belonging to one group.

37. For the subfields of all fields identified by TAGs out of such an interlinked group, the following conventions apply:

    (a) if there are two or more sequences of reoccurring subfields (i.e., subfields characterized by the same identifier) in fields belonging to a group of interlinked fields, the number of repetitions n must be the same for all reoccurring subfields;

    (b) if m is the ordering number of a reoccurring subfield in the sequence of repetitions, those and only those subfields having the same ordering number m are considered as linked.

Survey of standardized tags

38. The following TAGs are used for identifying the recorded information:

TAG	DEFINITION	REMARKS
110	Number of document	
120	Plain language designation of kind of document	
131	WIPO Kind-of-Document Code ST.16	
132	Information further specifying ST.16	
151	Identification of source furnishing the record and owning the copyright	
190	Code of publishing country or organization	WIPO Standard ST.3

39. The following TAGs identify domestic filing data of the industrial property right concerned:

TAG	DEFINITION	REMARKS
210[1]	Number assigned to the (domestic) application	
220[1]	Filing date of the (domestic) application	ISO 8601

221[1]	Kind-of-Application Code	no corresponding WIPO Standard exists so far
231	Exhibition filing date	(same format as 220)
232[1]	Filing date of a complete specification	(same format as 220)
240	Date from which industrial property rights may have effect	(same format as 220)
250[1]	Language of original filing of (domestic) application	ISO R 639
260[1]	Language in which the (domestic) application was published	ISO R 639

[1]: Linked TAGs (see paragraphs 7 and 8, above).

40. The following TAGs identify priority data and patent family information:

TAG	DEFINITION	REMARKS
310[2]	Number(s) assigned to the priority application	(same format as 220)
320[2]	Date(s) of filing of priority application	WIPO Standard ST.3
330[2]	Code(s) of country(-ies) or organization(s) in which the priority application(s) was(were) filed	
350	Patent family information	

[2]: Linked TAGs (see paragraphs 7 and 8, above).

41. The following TAGs identify dates concerned with making available to the public the Industrial Property Right concerned:

TAG	DEFINITION	REMARKS
410	Date of making available to the public by viewing, or copying on request, an *unexamined* document, on which no grant has taken place on or before the said date	(same format as 220)
420	Date of making available to the public by viewing, or copying on request, an *examined* document, on which no grant has taken place on or before the said date	(same format as 220)
430	Date of publication by printing or similar process of an *unexamined* document, on which no grant has taken place on or before the said date	(same format as 220)
440	Date of publication by printing or similar process of an *examined* document, on which no grant has taken place on or before the said date	(same format as 220)
450	Date of publication by printing or similar process of a document on which grant has taken place on or before the said date	(same format as 220)
460	Date of publication by printing or similar process of the claim(s) only of a document	(same format as 220)
470	Date of making available to the public by viewing, or copying on request, a document on which grant has taken place on or before the said date	(same format as 220)

42. The following TAGs identify technical subject matter information:

TAG	DEFINITION	REMARKS
510	International Patent Classification, unspecified	For all TAGs on IPC a subfield giving the IPC version has to be foreseen
511	International Patent Classification, main (first) invention symbol	
512	International Patent Classification, further invention symbol(s)	
513	International Patent Classification, classification symbol(s) representing additional information	
514	International Patent Classification, linked indexing codes	
515	International Patent Classification, unlinked indexing code(s)	

520	Symbol(s) of domestic or national classification, unspecified	All TAGs on national or domestic classifications should also contain a subfield giving the country code of the originator of the respective classification
522	Main (first) symbol of national or domestic classification	
523	Further symbol(s) of national or domestic classification	
524	Linked indexing symbols of national or domestic classification	
525	Unlinked indexing symbols of national or domestic classification	
530	Universal Decimal Classification	
540[3]	Language of title of the invention	ISO R 639
541[3]	Title of the invention	
550[4]	Specification of keyword or descriptor language	
551[4]	Keywords, descriptors	
560	List of prior art documents, unstandardized	
561	Prior art patent document cited, standardized	
562	Prior art non-patent document cited	
570[5]	Language of abstract	ISO R 639
571[5]	Text of abstract	
572[6]	Language of claim(s)	
573[6]	Text of claim(s)	
580	Field of search, unstandardized	
581	Field of search in terms of IPC	
590[7]	Language of full text of patent document	ISO R 639
591[7]	Full text of patent document	
592[7]	Number of pages of patent document	

[3, 4, 5, 6, 7]: Linked TAGs (see paragraphs 7 and 8, above).

43. The following TAGs identify references to other legally related domestic patent documents including unpublished applications therefor:

TAG	DEFINITION	REMARKS
610[8]	Number of the earlier application to which the present document is an addition	
611[8]	Filing date of the earlier application to which the present document is an addition	(same format as 220)
612[8]	Kind-of-Application code of the earlier application to which the present document is an addition	(see TAG 221)
613[9]	Number of the earlier publication to which the present document is an addition	
614[9]	Date of the earlier publication to which the present document is an addition	(same format as 220)
615[9]	Kind-of-Document code of the earlier publication to which the present document is an addition	(see TAG 131)
620[10]	Number of the earlier application from which the present document has been divided out	
621[10]	Filing date of the earlier application from which the present document has been divided out	(same format as 220)
622[10]	Kind-of-Application code of the earlier application from which the present document has been	(see TAG 221)

	divided out	
630[11]	Number of the earlier application of which the present document is a continuation	
631[11]	Filing date of the earlier application of which the present document is a continuation	(same format as 220)
632[11]	Kind-of-Application code of the earlier application of which the present document is a continuation	(see TAG 221)
640	Number of the earlier publication which is “reissued” by the document concerned	
641	Kind-of-Document code of the earlier publication which is “reissued” by the document concerned	(see TAG 131)
650[12]	Number of a previously published patent document concerning the same application	
651[12]	Kind-of-Document code of the previously published patent document concerning the same application	(see TAG 131)

[8, 9, 10, 11, 12]: Linked TAGs (see paragraphs 7 and 8, above).

44. The following TAGs identify parties concerned with the patent document described by the record:

TAG	DEFINITION	REMARKS
710[13]	Name of applicant, unspecified	
711[13]	Name of individual applicant	
712[13]	Name of collective applicant	
713[13]	Residence (address) of applicant	
714[13]	WIPO Country code ST.3 of residence of applicant	
715[13]	WIPO Country Code ST.3 of nationality of applicant (if natural person)	
720[14]	Name of inventor	
721[14]	Residence (address) of inventor	
722[14]	WIPO Country Code ST.3 of residence of inventor	
723[14]	WIPO Country Code ST.3 of nationality of inventor	
730[15]	Name of grantee	
731[15]	Address of grantee	
732[15]	WIPO Country Code ST.3 of residence of grantee	
740[16]	Name of attorney or agent	
741[16]	Address of attorney or agent	
742[16]	WIPO Country Code ST.3 of residence of attorney or agent	
750[17]	Name of inventor who is also applicant	To be used preferably for countries where this identity is required by the law
751[17]	Address of inventor who is also applicant	
752[17]	WIPO Country Code ST.3 of nationality of inventor who is also applicant	
760[18]	Name of inventor who is also applicant and grantee	(see TAG 750)
761[18]	Address of inventor who is also applicant and grantee	
762[18]	WIPO Country Code ST.3 of nationality of inventor who is also applicant and grantee	

[13, 14, 15, 16, 17, 18]: Linked TAGs (see paragraphs 7 and 8, above).

45. The following TAGs identify data related to international conventions other than the Paris Convention:

TAG	DEFINITION	REMARKS
810	List of designated States according to the PCT, Chapter 1	WIPO Country Code ST.3 to be used, together with kind-of-application code
820	List of elected States according to the PCT, Chapter 2	(see TAG 810)
840	List of designated Contracting States under regional patent treaties	(see TAG 810)
850[19]	Date of fulfillment of the requirements of Article 22 and/or 39 of the PCT for introducing the national procedure according to the PCT	(same format as 220)
860[19]	Number of the international or regional application	
861[19]	Filing date of the international or regional application	(same format as 220)
862[19]	Language in which the international or regional application was originally filed	ISO R 639
870	Number of the international or regional publication	
871	Publication date of the international or regional publication	(same format as 220)
872	Language(s) of the international or regional publication	ISO R 639
880	Date of deferred publication of the search report	(same format as 220)
890	Document number of the original document according to the Havana Agreement	
891	WIPO Kind-of-Document Code ST.16 of the original document according to the Havana Agreement	
892	Country of origin of the original document according to the Havana Agreement	(see TAG 810)
893	Date of recognition of the Industrial Property Right according to the Havana Agreement	(same format as 220)

[19]: Linked TAGs (see paragraphs 7 and 8, above).

[End of Standard]