

## USA STANDARD FOR A FORMAT FOR BIBLIOGRAPHIC INFORMATION INTERCHANGE ON MAGNETIC TAPE

*The Chairman of the United States of America Standards Institute, Sectional Committee Z39, Library Work and Documentation, has approved publication of the following draft "USA Standard for a Format for Bibliographic Information Interchange on Magnetic Tape" to hasten availability of this fundamental contribution to bibliographic standardization. Two important implementations follow the Standard. Part B of Appendix I is "Preliminary Guidelines for the Library of Congress, National Library of Medicine, and National Agricultural Library Implementation of the Proposed American Standard for a Format for Bibliographic Information Interchange on Magnetic Tape as Applied to Records Representing Monographic Materials in Textual Printed Form (Books)"—more succinctly known as MARC II. Part C is a Committee working paper entitled "Preliminary Committee on Scientific and Technical Information (COSATI) Guidelines for Implementation of the USA Standard."*

### 0. INTRODUCTION

- 0.1 This introduction is not part of the proposed standard but is included to facilitate its use.
- 0.2 This standard defines a format which is intended for the *interchange* of bibliographic records on magnetic tape. It has not been designed as a record format for retention within the files of any specific organization. Nor has it been the intent of the subcommittee to define the content of individual records. Rather it has attempted to describe a generalized structure which can be used to transmit between systems records describing all forms of material capable of bibliographic descriptions, as well as related records, such as authority records for authors and subject headings.

- 0.3 In designing the format the subcommittee has tried to achieve the goals listed below. It recognizes, however, that the goals were not completely compatible and that various trade-offs were required.
- (a) *Hospitality*—to all kinds of bibliographic information should be provided;
  - (b) *Hardware independence*—a format which can be used with a variety of digital computers should be defined;
  - (c) *Uniformity of structure*—the structure of all machine records should be basically identical and include such control information as may be required to specify unique characteristics. For any given class of records the components of the format may have specific meanings and unique characteristics;
  - (d) *Data Manipulation*—the methods of recording and identifying data should provide for maximum manipulability leading to ease of conversion to other formats for various uses.
- 0.4 The standard includes the concept that the bibliographic unit may be described independently or in relation to other bibliographic units. Many relationships exist, including: the hierarchical, in which the bibliographic unit contains, or is contained in, another bibliographic unit, e.g., a monograph in a series; the equivalent, e.g., a work and its translation; and the sequential, e.g., a serial which appeared under a succession of titles. The standard provides for bibliographic records which describe one or more related bibliographic units, and provides for coding the relationships among them. Appendix II describes a proposed method for implementing this concept.
- 0.5 Preliminary guidelines for implementing the standard by two different groups of users are provided in Appendix I. These guidelines are not part of the standard but are included to illustrate the use of the format.
- 0.6 Explanatory material which is not part of the standard but which will assist in its interpretation or implementation appears in brackets.
- 0.7 The appendices accompanying this standard are not part of the standard.
- 0.8 The development of this standard was made possible partially by support received from the National Science Foundation and the Council on Library Resources. Personnel of the USASI Committee Z39 at the time the Committee approved the standard were Dr. Jerrold Orne, Chairman; Mr. James Wood, Vice-Chairman; and Mr. Harold Oatfield, Secretary.

The Subcommittee on Machine Input Records, which is directly responsible for this standard, had the following personnel:

Mrs. Henriette D. Avram, Chairman  
Assistant Coordinator of Information Systems  
Information Systems Office  
Library of Congress  
Washington, D.C. 20540

Mrs. Pauline A. Atherton  
School of Library Science  
Syracuse University  
308 Carnegie Library  
Syracuse, New York 13210

Mr. Arthur R. Blum  
American Institute of Physics  
335 East 45th Street  
New York, New York 10017

Mr. Lawrence F. Buckland  
President, Inforonics, Inc.  
806 Massachusetts Avenue  
Cambridge, Massachusetts 02139

Miss Ann T. Curran  
Inforonics, Inc.  
806 Massachusetts Avenue  
Cambridge, Massachusetts 02139

Mr. Kay D. Guiles  
Information Systems Office  
Library of Congress  
Washington, D.C. 20540

Mr. Frederick G. Kilgour  
Director, Ohio College Library Center  
1314 Kinnear Road  
Columbus, Ohio 43212

Mr. Abraham I. Lebowitz  
Assistant to the Director  
National Agricultural Library  
U.S. Department of Agriculture  
Washington, D.C. 20250

Mrs. Phyllis B. Steckler  
R. R. Bowker Company  
1180 Avenue of the Americas  
New York, New York 10036



## 1. GLOSSARY

It has been considered unnecessary to define terms in common use. Terms which have a special meaning in the standard or which might be ambiguous are defined below.

**BASE ADDRESS OF DATA.** A data element whose value is equal to the character position of the character following the field terminator of the directory, where the specified origin is the first character of the leader. [Example: If the directory contains two (2) entries, the first character position of data will be 49, and therefore the base address of data equals 49.]

**BASIC CHARACTER.** A character occurring in columns 2, 3, 6 or 7 of the Standard Code as defined in USAS X3.4-1967 Code for Information Interchange, p. 6. [The basic character set is included as part of the illustration on page 82 of Appendix I, columns 2, 3, 6 and 7.]

**BIBLIOGRAPHIC INFORMATION INTERCHANGE FORMAT.** A format for the exchange, rather than the local processing, of bibliographic records. (The terms "bibliographic information interchange format," "information interchange format," and "interchange format" are used interchangeably in this standard.)

**BIBLIOGRAPHIC LEVEL.** A data element which, in conjunction with the data element "type-of-record," specifies the characteristics and describes the components of the bibliographic record. [See Appendix I for an illustration of an application of this data element.]

**BIBLIOGRAPHIC RECORD.** A collection of fields, including a leader, directory, and bibliographic data, describing one or more bibliographic units treated as one logical entity.

**BIBLIOGRAPHIC UNIT.** A defined body of recorded information and the artifact on which it is recorded, e.g., a book, chapter of a book, map, cuneiform tablet, digital magnetic tape file, song (sheet music), and song (phonograph record). A bibliographic unit may be part of a larger bibliographic unit (e.g., the chapter as part of a book, which in turn is part of a series). [It is assumed that the originators of bibliographic information and/or bibliographic descriptions follow a set of rules or guidelines which define, for the originating source, what is to be treated as a bibliographic unit.] A single author or subject heading authority record is also a bibliographic unit.

**CHARACTER.** *See* INTERNAL CHARACTER.

**COMMUNICATIONS FORMAT.** *See* BIBLIOGRAPHIC INFORMATION INTERCHANGE FORMAT.

**CONTROL FIELD.** A variable field which supplies parameters which may be required in the processing of the bibliographic record.

**CONTROL NUMBER.** An alphanumeric symbol uniquely associated with a bibliographic record assigned by the organization creating the bibliographic record.

**DATA ELEMENT.** A defined unit of information within a system.

**DATA ELEMENT IDENTIFIER.** A code consisting of one or more basic characters used to identify individual data elements within a variable field. If and when data element identifiers are used, each occurrence must be immediately preceded by a delimiter, and each data element identifier must immediately precede the data element it identifies. The length (in characters) of the data element identifier must be uniform for each field of a given record. [In effect, a delimiter and data element identifier are combined to form a symbol used to initiate and identify data elements within a variable field. The use of the concept of data element identifiers is optional and provides a means of explicitly identifying data elements, even though in some instances there may be a redundancy of identification (e.g., if a variable field consists of only one data element, presumably the tag alone would provide sufficient identification).]

**DATA FIELD.** A variable field containing bibliographic or other data not intended to supply parameters for the processing of the bibliographic record.

**DELIMITER.** A character which serves as an initiator, a separator, or a terminator of individual data elements within a variable field. [Whether a delimiter is used to initiate, to separate, or to terminate, is dependent upon a specific system.]

**DELIMITER (OR DELIMITER PLUS DATA ELEMENT IDENTIFIER) COUNT.** A data element whose value is the length (in characters) of the delimiter (or, if data element identifiers are used, the length (in characters) of the delimiter plus data element identifier) used within the record.

**DIRECTORY.** An index to the location of the variable fields (control and data) within a bibliographic record. The directory consists of entries.

**ENTRY.** A fixed field within the directory which contains information about a variable field.

**ENTRY MAP.** A data element which is used to indicate the structure of the entries in the directory.

**EXTERNAL CHARACTER.** A graphic symbol which may be represented by one or a series of two or more internal characters. [The external character "space" is always represented by an internal character.]

**FIELD.** A defined character string which may contain one or more data elements. *See also* CONTROL FIELD; DELIMITER; ENTRY; FIXED FIELD; INDICATOR; VARIABLE FIELD.

**FIELD TERMINATOR (FT).** A character used to terminate a variable field within a bibliographic record. The last variable field is terminated by a record terminator and not a field terminator.



- FILE.** A set of related records denoted by a single name.
- FIXED FIELD.** One in which every occurrence of the field has a length of the same fixed value regardless of changes in the contents of the field from occurrence to occurrence.
- FORMAT.** *See* STRUCTURE.
- FT.** *See* FIELD TERMINATOR.
- INDICATOR.** A data element associated with a data field which supplies additional information about the associated data field.
- INDICATOR COUNT.** A data element whose value is the length (in characters) of the indicator(s) which appears as the first data element in each variable data field. The length (in characters) of the indicator(s) must be uniform for each field of a given record. (A length of zero (0) is permitted).
- INFORMATION INTERCHANGE FORMAT.** *See* BIBLIOGRAPHIC INFORMATION INTERCHANGE FORMAT.
- INTERCHANGE FORMAT.** *See* BIBLIOGRAPHIC INFORMATION INTERCHANGE FORMAT.
- INTERNAL CHARACTER.** A pattern of bits of a predetermined length (depending on the system) treated as a meaningful unit. (The terms "internal character" and "character" are used interchangeably in this standard.)
- LEADER.** A fixed field which occurs at the beginning of each bibliographic record which provides parameters for the processing of the record.
- PADDING CHARACTER.** A character used to fill areas in fixed fields which contain no data. [*See* paragraph A.2.1.4 of Appendix I.]
- PRIMARY BIBLIOGRAPHIC UNIT.** That bibliographic unit whose physical and bibliographic characteristics determine the type-of-record and bibliographic level.
- RECORD.** *See* BIBLIOGRAPHIC RECORD.
- RECORD LENGTH.** A data element whose value is equal to the length (in characters) of the bibliographic record including the record terminator.
- RECORD TERMINATOR (RT).** A character used to terminate each record.
- RT.** *See* RECORD TERMINATOR.
- STATUS.** A data element which indicates the relation of the bibliographic record to a file (e.g., new, updated, etc.).
- STRUCTURE.** The framework of fixed and variable fields within the bibliographic record.
- SUBRECORD.** A group of fields within a bibliographic record which may be treated as a logical entity. [When a bibliographic record describes more than one bibliographic unit, the descriptions of individual bibliographic units may be treated as subrecords.]

**TAG.** A series of characters used to specify the name or label of an associated variable field.

**TYPE-OF-RECORD.** A data element which, in association with the data element "bibliographic level," indicates the form of the bibliographic description provided for the primary bibliographic unit. [It is assumed that the person providing the bibliographic description, on the basis of predefined criteria, will determine the treatment a given item is to receive; i.e., whether the item is to be treated as a book, a journal article, a map, a picture, an abstract, a bibliographical footnote, etc. If a given item consists of parts which, if they occurred independently, would be accorded different bibliographic descriptions, the choice of treatment selected is assumed to be the most appropriate. Frequently occurring combinations may be accorded their own treatments, e.g., collections of drawings with accompanying text. For each established form of bibliographic description, there will be a record format whose components are defined by the "type-of-record" data element. Among these components are the length of the fixed fields, the tagging scheme employed, and the definition of the data elements. If the interchange format is used for the interchange of records of a type for which "bibliographic description" is not a parameter, e.g., authority records, this data element may be redefined. See Appendix I for an illustration of an application of this data element.]

**VARIABLE FIELD.** One in which the length of an occurrence of the field is determined by the length (in characters) required to contain the data stored in that occurrence. The length may vary from one occurrence to the next.

## 2. PURPOSE AND SCOPE

- 2.1 This standard defines a format for the interchange of bibliographic and related [authority files, subject heading lists, etc.] records.
- 2.2 This standard does not define a record format for retention within the files of any specific organization.
- 2.3 This standard does not necessarily define the content of individual records. It does describe a generalized structure which can be used for the interchange of records describing various forms of bibliographic material.
- 2.4 This standard assumes the utilization of the following USASI Standards and Proposed Standards:
  - (a) USAS X3.22-1967 Recorded Magnetic Tape for Information Interchange (800 CPI, NRZI)
  - (b) USAS X3.4-1967 Code for Information Interchange
  - (c) Proposed Standard X3.2/552 Magnetic Tape Labels and File Structure



## 3. BIBLIOGRAPHIC INFORMATION INTERCHANGE FORMAT

3.1 *Schematic Representation*

The interchange format is schematically represented below:

Leader	Directory	F T	Control Number	F T	Other Control Fields (If Present)	F T	Data Field 1	F T	Data Field n	R T
--------	-----------	--------	-------------------	--------	---	--------	--------------------	--------	--------------------	--------

3.2 *Leader*3.2.1 *Schematic Representation*

The leader is schematically represented below:

RECORD LENGTH	STATUS	TYPE OF RECORD	BIBLIO- GRAPHIC LEVEL	RESERVED FOR FUTURE USE	INDI- CATOR COUNT	DELIMITER (OR DE- LIMITER PLUS DATA ELEMENT IDENTIFIER COUNT	BASE ADDRESS OF DATA	RESERVED FOR USE BY USER SYSTEMS	ENTRY MAP
0	4 5	6	7	8	9 10	11	12	16 17	19 20 23

3.2.2 *Record Length*

The record length is a 5-digit decimal number equal to the bibliographic record length. This number will include its own five characters and the record terminator. The record length will always be present in character positions 0-4 of the record. In the interchange format the bibliographic record has a maximum length of 99,999 characters.

3.2.3 *Status*

A data element in character position 5 consisting of 1 basic character.\*

3.2.4 *Type-of-Record*

A data element in character position 6 consisting of 1 basic character.\*

3.2.5 *Bibliographic Level*

A data element in character position 7 consisting of 1 basic character.\*

3.2.6 *Indicator Count*

A data element in character position 10 consisting of 1 decimal digit\* equal to the length (in characters) of the indicator(s) which appears as the first data element of each variable data field. If indicators are not used, this field is set to zero ( $\emptyset$ ). (See 3.4.2.1)

\*See Appendix I for an illustration of an application of this data element.



### 3.2.7 *Delimiter (or Delimiter Plus Data Element Identifier) Count*

A data element in character position 11 consisting of 1 decimal digit equal to the length (in characters) of the delimiter (or, if data element identifiers are used, the length (in characters) of the delimiter plus data element identifier) used within the record. If a delimiter is not used, this field is set to zero ( $\emptyset$ ). If a delimiter alone (i.e., without data element identifiers) is used, this field is set to one (1).

### 3.2.8 *Base Address of Data*

A data element in character positions 12-16 consisting of 5 decimal digits and equal to the combined length (in characters) of the leader and directory (including the field terminator at the end of the directory).

### 3.2.9 *Entry Map* (See 3.3.1 for the description of entries.) Structure of each entry in the directory:

Tag	Length of Field	Starting Character Position
-----	-----------------	-----------------------------

Entry map:

m	n	$\emptyset$	$\emptyset$
---	---	-------------	-------------

$m$  = length (in characters) of the "length of field" portion of each entry in the directory

$n$  = length (in characters) of the "starting character position" portion of each entry in the directory

$\emptyset$  = undefined; available for future use

The entry map is a data element in character positions 20-23 consisting of 4 decimal digits. Each decimal digit recorded corresponds sequentially to each portion of the entry, except for the portion allotted to the tag. Character position 20 in the entry map indicates the length (in characters) of the "length of field" portion of each entry in the directory; character position 21 indicates the length (in characters) of the "starting character position" portion of each entry. If one of these does not occur, the relevant character position in the entry

map is set to zero. Character positions 22 and 23 are undefined and are available for future use. [Since bibliographic data is usually variable in length, the structure of an entry in the directory will usually follow the pattern "tag, length of field, starting character position." The inclusion of an entry map provides flexibility for those users who wish to structure the entry in the directory differently, either by including (in addition to tag, length of field, and starting character position) other data elements not defined in this standard or by excluding those that have been defined. However, any restructuring of the entry by a user will have to be done within the general limitations imposed by the standard (see 3.3.1). The use of the entry map can be illustrated as follows: (1) An entry map set to 4500 would define the characteristics of a directory in which each entry consisted of a 3-digit tag (not expressed in the entry map), a 4-digit length of field, and a 5-digit starting character position. (2) An entry map set to 0500 would define the characteristics of a directory in which each entry consisted of a 3-digit tag, no length of field data element, and a 5-digit starting character position. See Appendix I for an illustration of an actual application of the concept of an entry map.]

### 3.3 *Directory*

The directory consists of a series of fixed fields (hereinafter referred to as entries). The directory ends with a field terminator. The directory must contain at least one entry for each subsequent variable field (control and data). [In the case of very long fields additional entries may be required. See 3.3.1.3.]

#### 3.3.1 *Entries*

Each entry consists of 12 characters. Each entry must contain, at the very least, a tag, and length of field, or a tag and starting character position and must correspond, unambiguously, to a specific variable length data or control field. The tag, length of field, and starting character position must, whenever they occur, be in that sequence.

##### 3.3.1.1 *Tag*

The tag is a data element consisting of 3 basic characters.

##### 3.3.1.2 *Tags for Control Fields*

Tags 001-009 are reserved for control fields as shown:



001	Control number
002	Reserved for Subrecord directory, if any*
003	Reserved for Subrecord relationship, if any*
004-009	Reserved for use by user systems

### 3.3.1.3 *Length of Field*

The length of field in the entry is the length (in characters) of the variable field to which it corresponds. The length of field includes the indicator(s) and field terminator. It is expressed as a decimal number.

If the length of a variable field exceeds the maximum length expressible as decimal number in the length of field portion of the entry, two or more entries (called a "subset" for the purposes of this explanation) will be used to define the location and extent of such a field. Since all the entries in the subset of entries reference the same variable field, they will contain the same tag. The length of field in each entry of the subset, except the last entry in the subset, will be set to  $\emptyset$  to indicate that the length of field is equal to the maximum length expressible and that there is additional information for the same field in the next entry in the record directory. The length of field for all entries in the subset subsequent to the first will refer to the length (in characters) of the overflow data. [This convention cannot be followed if the structure of the entry does not contain a length of field.]

### 3.3.1.4 *Starting Character Position*

The starting character position is the character position of the first character in the variable field (which may be an indicator or data; see 3.4.2) referenced by the entry. It is given relative to the base address of data (i.e., the first character of the first variable field following the directory is numbered  $\emptyset$ ).

### 3.3.2 *Sequence of Entries*

The entries in the directory may be recorded in any sequence (i.e., they need not be in the same sequence as the corresponding variable fields) except that the

\*Appendix II illustrates a possible method of handling subrecords within a bibliographic record. This is not part of the Standard.

entry for tags 001-009 must always be first and in ascending numeric sequence. [Note that specific systems may use the sequence of entries in the directory to convey semantic information.]

### 3.4 *Variable Fields*

#### 3.4.1 *General*

Following the leader and directory, the bibliographic record consists of variable fields. (Although the directory is technically a variable field, the following paragraphs do not apply to it.)

#### 3.4.2 *Structure of Variable Fields*

Each variable field consists of indicators(s) (if used), a delimiter (if used), a data element identifier (if used), data, and a field terminator, as shown. Control fields do not contain indicators, delimiters, or data element identifiers.

INDICATOR(S)*	DELIMITER*	DATA ELEMENT IDENTIFIER*	DATA	FIELD TERMINATOR
---------------	------------	--------------------------	------	------------------

\* Except control fields

##### 3.4.2.1 *Indicator*

The indicator is the first data element in each variable field. The length (in characters) of the indicator(s), which may be  $\emptyset$ , (i.e., no indicator is present) is recorded in the indicator count in the leader. All variable fields, except control fields, in the same record have the same length (in characters) for an indicator(s).

##### 3.4.3 *Sequence of Variable Fields*

The variable fields, except for the control fields associated with tags 001-009, need not occur in the same sequence as the corresponding directory entries. The control fields which occur must be first and in ascending numeric sequence.

##### 3.4.4 *Control Fields*

The variable fields associated with tags 001-009 are control fields. Control fields do not contain indicators, delimiters, or data element identifiers.

###### 3.4.4.1 *Control Number Field*

This field contains the control number, consisting of basic characters. This field must always occur once, and only once, in each



bibliographic record, and must immediately follow the directory.

### 3.5 *Variable Data Fields*

#### 3.5.1 *General*

The remainder of the bibliographic record consists of variable data fields. There are no restrictions on the number, length, or content of the variable data fields other than those already stated or implied (e.g., those based on the limitations of the total record length).

#### 3.5.2 *Multiple Data Elements*

Multiple data elements within fields may be fixed or variable and may be identified by position, by the use of a delimiter alone, or by the use of a delimiter plus data element identifier(s) as the case may be.