DOCUMENT RESUME

ED 104 436 IR 001 840

AUTHOR	Avram, Henriette D.
TITLE	Report on the IFLA Working Group on Content Designators.
PUB DATE	Nov 74
NOTE	18p.; Paper presented at the General Council Meeting of the International Federation of Library Associations (40th, Washington, D.C., November 16-23, 1974)
EDRS PRICE	MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS	*Bibliographic Citations; Cataloging; *Data Bases; *Information Storage; International Organizations; *International Programs
IDENTIFIERS	*Content Designators; IFLA 74; International Federation of Library Associations

ABSTRACT

In order to facilitate the international exchange of machine-readable bibliographic descriptions, the International Federation of Library Associations Working Group on Content Designators was organized in 1973 to arrive at a standard means of identifying data elements or providing additional information about a data element. In three meetings the group considered 52 working papers and arrived at 21 decisions, representing the majority viewpoint of the members. It was decided that the format would be designed to handle all media, including single-level and multi-level structures. Basic principles for assigning tags, indicators, and data element identifiers were also adopted. The 21 decisions are listed in concise form in Appendix B. (PF)

REPORT ON THE IFLA WORKING GROUP ON CONTENT DESIGNATORS

Prepared By

Henriette D. Avram
Library of Congress

July 1974

For presentation to the

IFLA Committee on Mechanization
Washington, D.C.
November 1974

BACKGROUND

The first published draft of the International Standard Bibliographic Description for monographs, issued in 1971, provided for a standard set of descriptive elements in a standard order using standard punctuation to separate the elements. The international standard format for information interchange in machine-readable form, ISO[1/]2709, published in 1973, provided for a standard format structure, the carrier of the bibliographic description for all forms of material. The ever-increasing number of national MARC programs provided the impetus to exchange machine-readable data across national boundaries. However, international MARC was not a reality. While the majority of national agencies adopted the ISBD and used the standard ISO 2709 format structure, there were differences in the identification of data elements as well as in the content of the records. Consequently, the need exists for each national agency to provide tailor-made procedures to process the records of every other national agency.

A format for bibliographic data is composed of three elements:

1. The structure of the record, which is the physical representation of the information on the machine-readable medium.

2. The content designators, which are the means of identifying data elements or providing additional information about a data element.

3. The content of the record, which is the data itself.

Content designation is dependent on cataloging codes and cataloging practices. Unfortunately, the cataloging codes of various countries still differ with respect to the choice and form of data. Consequently, this difference affects the content designation of the MARC records created by national agencies.

1/ International Organization for Standardization.

Recognizing the difficulties in exchanging machine-readable data created by this lack of standardization, the International Federation of Library Associations (IFLA) in 1973 established, under the auspices of the Committees on Cataloguing and Mechanization, an international Working Group on Content Designators. Appendix A lists the names and organizations of the members of the Working Group.

The Working Group has met three times: in Grenoble, France, August 1973; in Brussels, Belgium, February 1974; and in Helsinki, Finland, May 1974. Its goal is to publish a draft international MARC record for book materials at least by the fall of 1975 for comment by national bibliographic agencies. Fifty-two working papers have been written by the Working Group members to date. This paper reports progress through July 1974 and borrows heavily from the working papers.

SCOPE

The scope of responsibility of the IFLA Working Group is to arrive at a standard set of content designators for different forms of material for the international interchange of bibliographic data in machine-readable form. Within this scope, the Working Group will be concerned with the requirements for the international library community, i.e., libraries and national bibliographies. The magnitude of the assignment is such that it appeared unwise to include the needs of other bibliographic agencies such as abstracting and indexing services concurrently. However, liaison has been established with an ISO counterpart, the Working Group on Content Designators, ISO TC/46/SC4/WG2,[2/]

2/ The ISO Working Group has not been active to date pending the completion of the work of the IFLA Working Group.

and all working papers have been submitted to the chairman of that group. At the last meeting of the ISO Working Group in Helsinki, Finland, May 1974, the IFLA work was reviewed and a recommendation made that an ISO member serve as liaison on the IFLA Working Group to represent documentalists.

FUNCTIONS OF CONTENT DESIGNATORS

In order to begin discussions, it was essential that the Working Group agree on the function of content designators in general and on the definition and function of each content designator (tags, indicators, and data element identifiers) in particular.

Bibliographic data in machine-readable form permits great flexibility and allows the creation of a wide variety of products. To operate on data in an automated mode, four basic operations are performed.

1. Store - the storage operation is the machine management of the data base, i.e., the organization of the bibliographic files as well as the keys selected for access to the bibliographic records.

2. Retrieve - the retrieval operation is the selection of data from the data base. Retrieval includes the selection of a single element from a record; the selection of a single record, e.g., by a unique control number; and the selection of a category of records, e.g., all French-language monographs on a particular subject with an imprint date of 1968.

3. Arrange - the arrange operation puts information into a desired sequence.

4. Display - the display operation is the preparation of data for the purpose of making information human-readable, e.g., display on a CRT, computer printout, or photocomposed output.

In order to perform these four basic operations, content designators are assigned to the data content of the record. The Working Group, therefore, agreed that the function of content designators is to provide unambiguously the means for the user to store, retrieve, arrange and display information in a variety of ways to satisfy his requirements. Building on this general statement of the function of content designation, the following definitions of tags, indicators, and data element identifiers were adopted.

1. A tag is a string of characters used to identify or name the main content of an associated data field. The designation of main content does not require that a data field contain all possible data elements all the time.

2. An indicator is a character associated with a tag to supply additional information about a data field or to give parameters for machine processing of the data field.

3. A data element identifier is a code consisting of one or more characters used to identify individual data elements within a data field. The data element identifier precedes the data element which it identifies.

PRINCIPLES OF FORMAT DESIGN

Prior to the first meeting of the Working Group, a matrix[3/] was prepared listing the content designators used in various national MARC formats. It became obvious from a study of the matrix that it would not be possible to arrive at an international standard without formulating a set of basic

3/ The matrix was based on a working paper entitled "Table 1. Comparative Table for MARC Formats" issued by the Canadian National Library.

principles upon which the international standard would be built. The six principles enumerated below were agreed to by the Working Group at its second meeting:

1. The international format should be designed to handle all media. The forms of material will initially be limited to those forms[4] where either experience has been gained in the conversion to machine-readable form or in-depth analysis has been performed to define the elements of information for the material.

2. The international format should accept single-level and multi-level structures.[5] There is a requirement to express the relationship of one bibliographic entity to another. This relationship may take the form of: a) a hierarchical relationship, i.e., works which are part of a large bibliographic entity such as a single volume of a multi-volume set, or b) a linear relationship, i.e., works which are related to other works such as a book in translation. The method of expressing linear relationships is straightforward and not subject to varying methods of noting the relationship by different agencies. A record will exist for each item described, e.g., the original work and the translation of the original, and a pointer will be provided in each record to link it to the other record. However, for expressing hierarchical relationships, there are at least two methods presently in use. One method is to place the information on the related work in a single field within the

4/ Books, serials, maps, films, manuscripts, music, and sound recordings.
5/ It was later discovered that all members of the Working Group did not have the same understanding of the meaning of single-level and multi-level records. This principle was discussed at great length at the Helsinki meeting and at this time (July 1974) the issue is not fully resolved.

record. For example, the different volumes of a multi-volume set may be carried in a contents field or the series to which an item belongs may be carried in a series field. This method has been termed a single level structure.

A multi-level structure makes use of subrecords, i.e., the description of the multi-volume set would make up one subrecord[6/], and a subrecord would be used for the description of each volume of the multi-volume set. Each subrecord would have a level identification, e.g., the multi-volume set would be level 1 and each volume described in the set, level 2.

Although there is no international agreement on a technique to incorporate level capabilities in the exchange format, the Working Group recognized that national agencies must maintain the authority to record cataloging data in a manner that reflects their cataloging practices. Consequently, there is a requirement for single-level structure where relationships are expressed as fields within the single level record and multi-level structures where relationships are expressed as subrecords within the multi-level record.

The concept can be further clarified by the following outline showing the principal parameters of each level type:

Single-level (a single record)

1 record identification number

1 title proper tag

No level identification

6/ A subrecord contains bibliographic data which, in most instances, is an independent unit of information, e.g., the description of each volume in a multi-volume monograph. Consequently, the data making up the subrecord has data elements content designated.

Multi-level (a single record)

1 record identification number

More than 1 title proper tag possible

Level identification

3. Tags should identify a field by type of entry as well as function by assigning specific values to the character positions. Assigning values to the characters of the tags provides the flexibility to derive more than one piece of information from the tag, e.g., one tag could be defined for principal author, personal name; another, for secondary author, personal name. The first character position of the tag would be assigned the intellectual responsibility of the author, the second character position, the type of author, i.e., personal or corporate. This technique would allow the retrieval of all personal names regardless of relationship to item cataloged.

4. Indicators should be tag dependent and used as consistently as possible across fields. Indicators should be tag dependent because they provide both descriptive and processing information about a data field. If the value assigned to an indicator is used consistently, where possible, across fields, then machine coding may be simplified to process different functional fields which contain the same type of entry, e.g., personal names used as primary and secondary authors.

5. Data element identifiers should be tag dependent but, as far as possible, common data elements should be identified by the same data element identifier across fields. Although conceptually, making data element identifiers tag independent is attractive, the limited number of alphabetic, numeric, and special characters could restrict the number of data elements to be uniquely

identified. However, the principle of identifying common data elements across fields by the same identifier as far as possible adds flexibility to the machine manipulation of data, e.g., processing like elements such as dates across fields.

6. The fields in a bibliographic record are primarily related to broad categories of information relating to "Subject of Document," "Description of Document," "Intellectual Responsibility for Document," "Identity of Document," etc. In a MARC record, the primary grouping of fields should be according to those fundamental categories. Since the majority of national formats are arranged by using the function as primary notation[7] and the type of entry as secondary notation, there appears to be no useful purpose in reversing this order.

In reporting the progress made by the Working Group, it is difficult to relate events chronologically because they seldom occured that way. Relating the work performed and the decisions made by this particular working group is even more difficult because of the amount of effort that has been expended in accomplishing this task. The distillation of 52 working papers and the rationale behind each decision, based on those working papers, is beyond the scope of a progress report. In the course of defining the six basic principles given above, many other issues were brought to light and decisions made. These decisions have been summarized in Appendix B.

7/ Notation as used here refers to the meaning and values assigned to the character positions of tags.

FUNCTIONAL BLOCS

Principle 6 gave the direction for the next effort of the Working Group. Various papers were written describing the content of the functional blocs and these papers were discussed at length at both the Brussels and the Helsinki meetings.

From the onset of the deliberations on the definition of the functional blocs of the format, it was agreed by all concerned that the ISBD[8/] prescriptions as to data elements to be included, the order of these elements, and the punctuation and spacing separating those elements, would be followed in all details in the design of the descriptive bloc for the international format.

The elements of the ISBD make up a base record, i.e., a record that stands alone without any additional information (such as established forms of names, subjects, etc.) as the unique description of an item. As discussions continued, it was recognized that, with the present lack of an internationally accepted cataloging code, with the intellectual difficulties inherent in the formulation and the use of subject terms and classification systems, and with the language problems associated with crossing national boundaries, the international format should consist of a standard base record (the ISBD) and standard identification numbers (ISBN, ISSN, etc.) and all other information (reflecting national cataloging codes and practices) should be exchanged as used in national formats. Present thinking assumes that the national information

8/ Although the Working Group is concerned with a generalized format for all forms of material, the participants agreed that work could be concentrated initially on those forms for which an ISBD had been derived, namely monographs and serials. As an ISBD was published for each additional form of material, the content designators for that material would be added to the generalized format.

would be assigned unique but broad content designators so that the information could be recognized by the recipient of the data. The rationale behind this decision is that, as long as the machine-readable data has been encoded and recorded, it may as well be exchanged, thus giving the recipient the option to use it or not, depending on his local needs.

Based on the above, the functional blocs and their contents were specified as given below:

1. Identification Bloc -- Includes those elements that uniquely identify the record (e.g., control number, ISBN, ISSN, etc.).

2. Descriptive Bloc -- Includes those areas now covered by the ISBD (i.e., title, edition, imprint, collation, series, notes.).[9/]

3. Intellectual Responsibility Bloc -- Includes persons and corporate bodies. These names will be identified as: Personal, Family, Corporate, or Meeting, and will be further identified as having Primary, Alternative, or Secondary Responsibility. (A code will be used to show the type of responsibility in more detail.) Otherwise, these fields will include the content designators and punctuation as supplied by the issuing agency.

4. Subject Bloc -- Includes both systematic and verbal subject identification (e.g., UDC, PRECIS, LC Subject Headings, etc.). Personal and corporate names used as subjects will appear in this bloc and will include the content designators and punctuation as supplied by the issuing agency.

5. Linking Entry Bloc -- Includes standard links to other records such as ISBN, ISSN, Key Title, etc. In addition (until international networks are formed) it may include the local system number of the related record.

6. Related Entry Bloc -- Includes links in textual form as established by the issuing agency to other records (e.g., series added entry, author/title added entry, etc.). This kind of relationship will be shown by a tag or by an indicator.

9/ Although the notes are recorded in the language of the country where the item is cataloged, several participants of the Working Group felt their agencies would use the notes in the language given and others felt the notes could be translated without expending intellectual effort on their contents.

7. Related Title Bloc -- Includes unifo.. ..es, variant titles appearing on the piece, translated titles, etc., to be used as access points. These titles are not linking.

8. Coded Information Bloc -- Includes fixed-length data (frequently codes) describing various aspects of the work.

Each functional bloc will be further subdivided into fields and within each bloc, a group of fields will be reserved for local information, which will not be included in the international exchange format. In addition, a field will be reserved for future use to record the location symbols of the libraries holding a specific title.[10/]

The developing international format can be considered a hybrid standard, i.e., part of the format will probably be standard throughout, e.g., the data elements, the ordering of the elements, the punctuation, the content designators (tags, indicators, and subfield codes). The remainder of the format will be assigned standard tags but the data elements, the indicators, data element identifiers and punctuation will reflect national use. The benefits resulting from the decision to design the international format in this manner are several:

1. It will be possible to exchange bibliographic data across national boundaries in the near term without waiting for international agreement on cataloging codes and subject systems.

2. The standard part of the record, i.e., the base record, can be effectively used for a variety of purposes.

3. The entire record can be augmented and/or modified for national use.

10/ This decision was based on work in progress in the United States. Although a field has been reserved, it is too early to know with certainty whether holdings information will be interchanged in the bibliographic record or in a record specifically designed for holdings data or whether, in fact, there will be any real need to exchange this information across national boundaries.

4. Regardless of the procedures installed locally for the use of the data, the cataloging and conversion costs at each national agency should be reduced.

MAJOR AREA REQUIRING RESOLUTION

Having determined that the Descriptive Bloc would be the principle bloc where complete international standardization would be required, attention was turned to the data element identification in that bloc. Positions on the Descriptive Bloc varied, ranging from complete content and complex data element identification, through a less complete content and simpler data element identification, to a proposal for no identification beyond the ISBD punctuation. These differing opinions result from many factors. Experience in the conversion of bibliographic data to machine-readable form caused some members to emphasize the high cost of encoding these records. Some members felt it was the responsibility of the national agency to include the identification of every element for every purpose regardless of cost.

The ISBD punctuation was considered by some to be sufficient identification for each agency to automatically bring the record to the level of content designation detail required by that national agency. Others considered this technique too difficult in terms of computer programs required.

Although the majority of the members of the Working Group appear to be tending toward a simpler method of identification, the issue was not resolved at the Helsinki meeting. Four positions, varying from the most complex (complete content and complex identification) to the simplest (ISBD punctuation only) were set forth in a Working Paper and the members have been asked

to vote. Based on the response which is expected by August 1974, the functional blocs will be subdivided into fields with content designators assigned for discussion purposes at the IFLA meeting in November 1974.

APPENDIX A

LIST OF MEMBERS AND ORGANIZATIONS

Mrs. Henriette D. Avram (Chairman)
Library of Congress

Mr. Edwin Buchinski
National Library of Canada

Mr. Marc Chauveinc
Bibliotheque Interuniversitaire
de Grenoble

Mr. Richard E. Coward
The British Library

Mr. Rainer Erzepky
Deutsche Bibliothek

Dr. Walter Lingenberg (liaison from Committee on Mechanization)
Technical University
Berlin, Germany

Mr. Joel Poncet
Bibliotheque Nationale

Mrs. Lucia J. Rather
Library of Congress (Secretary)

Dr. A. L. Van Wesemael (liaison from Committee on Cataloguing)
University Library--Utrecht
The Netherlands

Mr. Mogens Weitemeyer
Det Kongelige Bibliotek
Copenhagen, Denmark

APPENDIX B

SUMMARY OF DECISIONS

INTRODUCTION

This paper documents decisions made by the Working Group from its initiation through the meeting in Helsinki, May 1974. The decisions represent a majority viewpoint of the members of the Working Group after consideration of the many working papers dealing with the issues involved. Consequently, this paper will serve as a basic document from this point forward. The decisions include the original six basic principles of format design which are stated as decisions 1, 5, 6, 8, 10, and 14 below.

DECISIONS

1. The international format will be designed to handle all media.

2. The type of media will be carried as coded information and also as textual information in the collation statement.

3. Content designators will be defined for manuscript material but has been assigned a low priority.

4. In the generalized format, a data element required for one form of material will be shown across all forms of material where applicable, i.e., in the description of the type--specific format within the generalized format, only those data elements applicable will be included. Consequently, a tag or data element identifier associated with a tag, selected for a specific data element will not be used for another data element in a form of material where the first data element does not appear.

5. The international format will accept both single-level and multi-level structures.1/

6. Tags should identify a field by type of entry as well as function by assigning specific values to the character positions.

7. Tags may be both numeric and alphabetic. First assignment will be numeric values, expanded to alphabetic (lower case preferred) if required.

8. Indicators should be tag dependent and used as consistently as possible across all fields.

9. Indicators may be both numeric and alphabetic. First assignment will be numeric values, expanded to alphabetic (lower case preferred) if required.

1/ It is not clear at this time if the majority of the Working Group agreed with this decision. Therefore, members are being asked to vote on this issue again.

10. Data element identifiers will be tag dependent, but, as far as possible, common data elements will be identified by the same data element identifiers across fields.

11. Data element identifiers may be both numeric and alphabetic. First assignment will be alphab lower case preferred), expanded to numeric if required.

12. Data element identifiers will be used in place of ISBD punctuation in those cases where both are applicable.

13. Data element identifiers will be given values for identification rather than for file arrangement.

14. The fields in a bibliographic record are primarily related to broad categories of information relating to "Subject of Document," "Description of Document," "Intellectual Responsibility for Document," "Identity of Document," etc. In a MARC record the primary grouping of fields should be according to these fundamental categories.

15. Descriptive information carried in notes will not be used as access points.

16. Any type-specific format within the generalized format will contain the number of linking fields required for that form of material.

17. Cataloging Source will be included in the generalized format as a fixed length data element.

18. Cataloging data related to several variant issues of one bibliographic title may be carried in a single record.

19. The library location symbol indicating a library holds a title in its collection will not be included in the format at this time. However, a tag will be reserved for future inclusion of this data.

20. Reference data will not be included in the generalized format for the exchange of bibliographic data.

21. Coded information (common data elements across media types and media-specific data elements) will be identified positionally in a coded information field.