

# **MARCdocs**

**The MARC 21 Bibliographic Format Database**

Version 1.2

October 28, 2004

William E. Moen, PhD

Jason Thomale

## 1 Introduction and Purpose

MARCdocs, the MARC 21 Documentation Database, is a pilot effort aimed at structuring the textual documentation from the MARC 21 Format for Bibliographic Data into a relational database. Using a database approach for the authoritative MARC documentation provides new opportunities for various applications, including more efficient maintenance of the documentation, easier and quicker updates and changes, exporting selected data in XML, as well as an aid to research into the development and evolution of MARC, a learning/reference tool for those seeking to understand the MARC format, and a tool to help those devising new bibliographic structures.

The database was first envisioned by Dr. William Moen of the University of North Texas School of Library and Information Sciences as a tool for his ongoing research into metadata utilization and interoperability. Dr. Shawne Miksa, Penelope Bernardino, JungWon Yoon, and Jason Thomale have carried this idea forward to implementation.

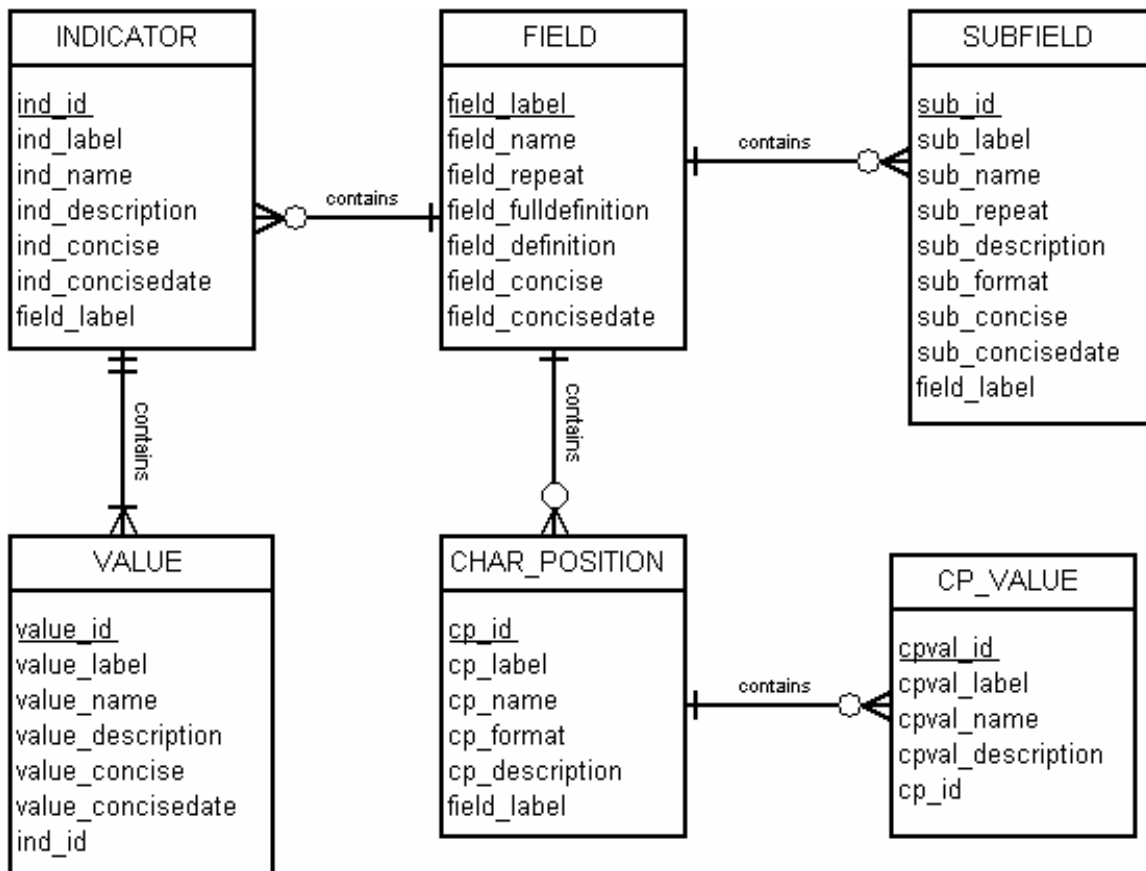
This application uses open source software tools including Linux, MySQL, and PHP. Jason Thomale, a masters student in the School of Library and Information Sciences, designed and implemented the database, developed procedures for data loading, and designed and programmed the web interface for the MySQL database.

## 2 Database Organization Overview

The MARCdocs data model mirrors the basic structure of most MARC fields as they are presented in the *MARC 21 Bibliographic Format* documentation. Six distinct entities relating to each field are represented:

1. Fields
2. Subfields
3. Indicators
4. Indicator Values
5. Character Position (for fields 006, 007, and 008)
6. Character Position Values (for fields 006, 007, and 008)

The Entity Relationship Diagram on the following page illustrates how these entities are related to one another:



### 3 Field

The Field entity stores general information about each field documented in the *MARC 21 Bibliographic Format* that can appear within a MARC record.

#### 3.1 Rules

Each field may have zero or more Subfields.

Each field may have zero or more Indicators.

Each field may have zero or more Character Positions.

#### 3.2 Data Structure

Data Element Name	Data Type	Description
FIELD_LABEL Primary Key	Char (3)	The 3-digit field code that uniquely identifies the field.
FIELD_NAME	Varchar (75)	The name of the field.
FIELD_REPEAT	Enum	Describes whether the field is repeatable or not. R=Repeatable, NR=Not Repeatable.
FIELD_FULLDEFINITION	Text	The field's definition as listed under "Field Definition and Scope" in the <i>MARC 21 Bibliographic Format</i> .
FIELD_DEFINITION	Text	The field's definition as listed in the Concise version of the <i>MARC 21 Bibliographic Format</i> .
FIELD_CONCISE	Enum	Describes whether the field appears in MARC 21 Concise—can be "Yes" or "No."
FIELD_CONCISEDATE	Varchar (4)	Describes the version of MARC 21 Concise from which data for this field was taken.

## 4 Subfield

The Subfield entity contains a listing of all possible subfields for each field.

### 4.1 Rules

Each subfield must belong to one and only one field.

### 4.2 Data Structure

Data Element Name	Data Type	Description
SUB_ID Primary Key	Varchar (7)	A 5-7 digit code that uniquely identifies the Subfield. Construct the SUB_ID as follows: parent field's FIELD_LABEL + S + SUB_LABEL. Subfield 'a' of field 010 would thus have a SUB_ID of 010Sa.
SUB_LABEL	Char (3)	The 3-digit subfield code.
SUB_NAME	Varchar (75)	The name of the subfield.
SUB_REPEAT	Enum	R=Repeatable, NR=Not Repeatable.
SUB_DESCRIPTION	Text	The subfield's description as listed under "Guidelines for Applying Content Designators" in the <i>MARC 21 Bibliographic Format</i> .
SUB_FORMAT	Set	The type of material to which this subfield applies. Can be one or more of the following: books, computer files, maps, music, continuing resources, visual materials, mixed materials, or sound recordings. If a subfield applies to all materials, use the word 'All.'
SUB_CONCISE	Enum	Describes whether the subfield appears in MARC 21 Concise—can be "Yes" or "No."
SUB_CONCISEDATE	Varchar (4)	Describes the version of MARC 21 Concise from which data for this subfield was taken.
FIELD_LABEL Foreign Key	Char (3)	A 3-digit code that uniquely identifies the parent field of the subfield.

## 5 Indicator

The Indicator entity lists all indicators belonging to a particular field.

### 5.1 Rules

Each indicator must belong to one and only one field.

Indicators may have one or more values.

### 5.2 Data Structure

Data Element Name	Data Type	Description
IND_ID Primary Key	Varchar (5)	A 5-digit code that uniquely identifies the Indicator. Construct the IND_ID as follows: parent field's FIELD_LABEL + I + IND_LABEL. The first indicator of field 022 would thus have an IND_ID of 022I1.
IND_LABEL	Enum	1 = First Indicator; 2 = Second Indicator
IND_NAME	Varchar (60)	The name of the indicator.
IND_DESCRIPTION	Text	The indicator's description as listed under "Guidelines for Applying Content Designators" in the <i>MARC 21 Bibliographic Format</i> .
IND_CONCISE	Enum	Describes whether the indicator appears in MARC 21 Concise—can be "Yes" or "No."
IND_CONCISEDATE	Varchar (4)	Describes the version of MARC 21 Concise from which data for this indicator was taken.
FIELD_LABEL Foreign Key	Char (3)	A 3-digit code that uniquely identifies the parent field of the indicator.

## 6 Value

The Value entity contains all possible values for each indicator.

### 6.1 Rules

Each value must belong to one and only one indicator.

‘Blank’ is an acceptable value. Always use ‘#’ to notate a blank.

### 6.2 Data Structure

Data Element Name	Data Type	Description
VALUE_ID Primary Key	Varchar (7)	A 7-digit code that uniquely identifies the value. Construct the VALUE_ID as follows: parent indicator’s parent field’s FIELD_LABEL + I + parent indicator’s IND_LABEL + V + VALUE_LABEL. Use ‘#’ for a blank value. The blank value of field 022’s first indicator would thus have a VALUE_ID of 022I1V#.
VALUE_LABEL	Char (1)	The 1-digit value code. Again, use ‘#’ for blank.
VALUE_NAME	Varchar (100)	The name of the value.
VALUE_DESCRIPTION	Text	The value’s description as listed under “Guidelines for Applying Content Designators” in the <i>MARC 21 Bibliographic Format</i> .
VALUE_CONCISE	Enum	Describes whether the indicator value appears in MARC 21 Concise—can be “Yes” or “No.”
VALUE_CONCISEDATE	Varchar (4)	Describes the version of MARC 21 Concise from which data for this indicator value was taken.
IND_ID Foreign Key	Varchar (5)	A 5-digit code that uniquely identifies the parent indicator of the value.

## 7 Char\_Position

For fields that contain lists of characters where the positions of the characters have significance (such as the 006, 007, and 008 fields), the Char\_Position entity keeps record of those characters' positions' meanings.

### 7.1 Rules

Each character position must belong to one and only one field.

The “2-letter format code” for building the CP\_ID varies depending on the character position's parent field. For fields 006 and 008, use the following (taken from the *MARC 21 Bibliographic Format*): Books = BK, Computer Files = CF, Maps = MP, Music = MU, Serials = SE, Visual Materials = VM, and Mixed Materials = MX; for All Materials, use AM. For the 007 field, use the following: MP = Map, CF = Computer File, GL = Globe, TA = Tactile Material, PG = Projected graphic, MI = Microform, NG = Nonprojected Graphic, MO = Motion Picture, KI = Kit, NO = Notated music, RE = Remote-sensing image, SO = Sound recording, TE = Text, VI = Videorecording, and UN = Unspecified; use AM for All Materials.

### 7.2 Data Structure

Data Element Name	Data Type	Description
CP_ID Primary Key	Varchar (9)	A 9-digit code that uniquely identifies the character position. Construct the CP_ID as follows: parent field's FIELD_LABEL + CP + 1 <sup>st</sup> applicable character position (CP_LABEL) + 2-letter format code.
CP_LABEL	Varchar (60)	The actual numerical character position(s). If listing more than one, separate by commas (i.e., 18,19,20,21).
CP_NAME	Varchar (60)	The name of the character position.
CP_FORMAT	Set	The format(s) to which the character position applies.
CP_DESCRIPTION	Text	The character position's description as listed under “Guidelines for Applying Content Designators” in the <i>MARC 21 Bibliographic Format</i> .
FIELD_LABEL Foreign Key	Char (3)	A 3-digit code that uniquely identifies the parent field of the character position.

## 8 CP\_Value

For fields that contain lists of characters where the positions of the characters have significance (such as the 006, 007, and 008 fields), the CP\_Value entity keeps record of those characters' positions' values' meanings.

### 8.1 Rules

Each character position value must belong to one and only one character position.

If a value is blank, use '#' in the CPVAL\_ID and the CPVAL\_LABEL.

### 8.2 Data Structure

Data Element Name	Data Type	Description
CPVAL_ID Primary Key	Varchar (16)	A 10-digit code that uniquely identifies the character position value. Construct the CPVAL_ID as follows: CP_ID + CPVAL_LABEL.
CPVAL_LABEL	Char (7)	The actual value.
CPVAL_NAME	Varchar (150)	The name of the character position value.
CPVAL_DESCRIPTION	Text	The value's description as listed under "Guidelines for Applying Content Designators" in the <i>MARC 21 Bibliographic Format</i> .
CP_ID Foreign Key	Varchar (9)	A 9-digit code that uniquely identifies the parent field of the character position.

## 9 Sample Queries

Note: For these sample queries, only a small portion of the total MARC Bibliographic Format has been entered into the database. Only the contents of fields 022, 045, and 018 exist. For the character position queries (section 9.4), the entirety of the database is used.

### 9.1 Field/Subfield Queries

#### QUERY 1

What subfields exist for the 045 field?

```
SQL: SELECT Field.field_label, Field.field_name, Subfield.sub_id,
       Subfield.sub_label, Subfield.sub_name
FROM Field JOIN Subfield ON Field.field_label =
       Subfield.field_label
WHERE Field.field_label = 045;
```

field_label	field_name	sub_id	sub_label	sub_name
045	Time Period of Content	045Sa	a	Time period code
045	Time Period of Content	045Sb	b	Formatted 9999 B.C. through C.E. time period
045	Time Period of Content	045Sc	c	Formatted pre-9999 B.C. time period
045	Time Period of Content	045S6	6	Linkage
045	Time Period of Content	045S8	8	Field link and sequence number

## QUERY 2

Which fields have Linkage subfields, and what are those subfields?

```
SQL: SELECT Field.field_label, Subfield.sub_label, Subfield.sub_name
FROM Field JOIN Subfield ON Field.field_label =
       Subfield.field_label
WHERE Subfield.sub_name LIKE '%link%';
```

field_label	sub_label	sub_name
022	6	Linkage
022	8	Field link and sequence number
045	6	Linkage
045	8	Field link and sequence number

## 9.2 Field/Subfield Format Queries

### QUERY 1

What formats apply to the 018 field?

```
SQL: SELECT DISTINCT Subfield.sub_format AS format, Field.field_label,
```

<b>field_label</b>	<b>field_name</b>	<b>sub_label</b>	<b>sub_name</b>	<b>sub_format</b>
022	International Standard Serial Number	a	International Standard Serial Number	Serials
022	International Standard Serial Number	y	Incorrect ISSN	Serials
022	International Standard Serial Number	z	Canceled ISSN	Serials
022	International Standard Serial Number	6	Linkage	Serials
022	International Standard Serial Number	8	Field link and sequence number	Serials
045	Time Period of Content	a	Time period code	All
045	Time Period of Content	b	Formatted 9999 B.C. through C.E. time period	All
045	Time Period of Content	c	Formatted pre-9999 B.C. time period	All
045	Time Period of Content	6	Linkage	All
045	Time Period of Content	8	Field link and sequence number	All
018	Copyright Article-Fee Code	a	Copyright article-fee code	Books,Serials

```
Field.field_name
FROM Field JOIN Subfield ON Field.field_label =
Subfield.field_label
WHERE Field.field_label = 018;
```

<b>format</b>	<b>field_label</b>	<b>field_name</b>
Books,Serials	018	Copyright Article-Fee Code

## QUERY 2

To what fields (and their subfields) does the “Serials” format apply?

```
SQL: SELECT Field.field_label, Field.field_name, Subfield.sub_label,
Subfield.sub_name, Subfield.sub_format
FROM Field JOIN Subfield ON Field.field_label =
Subfield.field_label
WHERE Subfield.sub_format LIKE '%Serials%' OR Subfield.sub_format
LIKE '%All%';
```

### 9.3 Field/Indicator/Value Queries

#### QUERY 1

What does it mean if the 022 field's first indicator is blank?

```
SQL: SELECT Field.field_label, Indicator.ind_label,
       Indicator.ind_name, Value.value_label, Value.value_description
FROM Field JOIN Indicator ON Field.field_label =
       Indicator.field_label JOIN Value ON Indicator.ind_id =
       Value.ind_id
WHERE Field.field_label = 022 AND Indicator.ind_label = 1 AND
       Value.value_label = '#';
```

field_label	ind_label	ind_name	value_label	value_description
022	1	Level of international interest	#	Value # indicates that the level of international interest is unknown or not specified. This value is used by all institutions other than the National Serials Data Program (NSDP) and ISSN Canada when recording the ISSN from an issue or from a bibliography.

## QUERY 2

What fields have indicators defined, and what are those indicators?

```
SQL: SELECT Field.field_label, Field.field_name, Indicator.ind_label,
       Indicator.ind_name
FROM Field JOIN Indicator ON Field.field_label =
       Indicator.field_label;
```

field_label	field_name	ind_label	ind_name
022	International Standard Serial Number	1	Level of international interest
045	Time Period of Content	1	Type of time period in subfield b or c

## 9.4 Character Position Queries

### QUERY 1

What does character position 18 in field 008 signify for the Music format?

```
SQL: SELECT Field.field_label, Field.field_name,
       Char_Position.cp_label,
       Char_Position.cp_name, Char_Position.cp_format,
       Char_Position.cp_description
FROM Field LEFT JOIN Char_Position ON Field.field_label =
       Char_Position.field_label
WHERE Field.field_label = 008 AND Char_Position.cp_label LIKE
       "%18%" AND Char_Position.cp_format = "Music";
```

<b>field_label</b>	<b>field_name</b>	<b>cp_label</b>	<b>cp_name</b>	<b>cp_format</b>	<b>cp_description</b>
008	FIXED- LENGTH DATA ELEMENTS	18,19	Form of composition	Music	A two-character code that indicates the form of composition. The codes are based on Library of Congress subject headings. If more than one code is appropriate, the code /mu/ (Multiple forms) is used in 008/18-19 and all appropriate specific codes are given in field 047 (Form of Composition).

## QUERY 2

For the above results, what are the possible values for this character position, and what do they mean?

```
SQL: SELECT CP_Value.cpval_label, CP_Value.cpval_name
FROM Field LEFT JOIN Char_Position ON Field.field_label =
Char_Position.field_label LEFT JOIN CP_Value ON
Char_Position.cp_id = CP_Value.cp_id
WHERE Field.field_label = 008 AND Char_Position.cp_label LIKE
"%18%" AND Char_Position.cp_format = "Music";
```

(PLEASE NOTE: Only the first 30 results are shown.)

<b>pval_label</b>	<b>cpval_name</b>
an	Anthems
bd	Ballads
bg	Bluegrass music
bl	Blues
bt	Ballets
ca	Chaconnes
cb	Chants, Other religions
cc	Chant, Christian
cg	Concerti grossi
ch	Chorales
cl	Chorale preludes

cn	Canons and rounds
co	Concertos
cp	Chansons, polyphonic
cr	Carols
cs	Chance compositions
ct	Cantatas
cy	Country music
cz	Canzonas
dg	Dance forms
dv	Divertimentos, serenades, cassations, divertisseme...
fg	Fugues
fm	Folk music
ft	Fantasias
gm	Gospel music
hy	Hymns
jz	Jazz
mc	Musical revues and comedies
md	Madrigals
mi	Minuets