CATALOGING & CLASSIFICATION QUARTERLY
https://doi.org/10.1080/01639374.2023.2257691

Consistency, Extent, and Validation of the Utilization of the MARC 21 Bibliographic Standard in the College Libraries of Assam in India

Bidyut Bikash Boruah, S. Ravikumar, and Fullstar Lamin Gayang

Department of Library and Information Science, North-Eastern Hill University, Shillong, India

ABSTRACT

This paper brings light to the existing practice of cataloging in the college libraries of Assam in terms of utilizing the MARC 21 standard and its structure, i.e., the tags, subfield codes, and indicators. Catalog records from six college libraries are collected and a survey is conducted to understand the local users' information requirements for the catalog. Places, where libraries have scope to improve and which divisions of tags could be more helpful for them in information retrieval, are identified and suggested. This study fulfilled the need for local-level assessment of the catalogs.

ARTICLE HISTORY
Received May 2023
Revised July 2023
Accepted September 2023

KEYWORDS
Cataloging; quality assessment; college libraries; library catalogs; MARC 21; MARC tags; catalogs of Assam

Introduction

Cataloging is a primordial concept that has been associated with libraries from the very beginning. The primary purpose is to describe a document so that, from the bibliographic description, the document can easily be retrieved from millions of documents in a collection.[1] The change in information literacy, information-seeking behavior of the users, modes of publication of the literature, and development of information and communication technology (ICT) have brought some changes and progression to the techniques, modes, and the process of cataloging from time to time.

The computers and integrated library management software use the machine-readable cataloging data format, MARC 21. It is a structured format that enables standard bibliographic records of books and other documents to be created, maintained, and manipulated by computer in a standard way. Nowadays, library catalogs are available through an OPAC (Online Public Access Catalog) or Web-OPAC with user-friendly interfaces. Keyword searching, natural word searching, etc., are now strengthening the user-friendly interfaces of the OPAC.

CONTACT S. Ravikumar ravikumarnehu@gmail.com Department of Library and Information Science, North-Eastern Hill University, Shillong, India

Though MARC 21 records can use fields from 001 to 999, libraries do not use all the tags. In the MARC 21 record, 10% of the tags are used repeatedly, and the other 90% are seen only occasionally or rarely.[2] "Understanding MARC: Bibliographic Machine-Readable Cataloging," published by the Network Development and MARC Standards Office at the Library of Congress, lists libraries' most frequently used MARC 21 tags in bibliographic records as 010, 020, 040, 100, 130, 240, 245, 246, 250, 260, 300, 440, 490, 500, 504, 505, 520, 600, 610, 650, 651, 700, 710, 740, 800, and 830.

This paper intends to study the utilization of the MARC 21 bibliographic standard in terms of extent and MarcEdit validation in the selected college libraries of Assam.

Literature review

Cataloging is a core function of a library. It is the foundation for all other library services or activities.[1] The mission of cataloging has been to facilitate easy retrieval by its users. MARC is accepted and used worldwide and is the basis for almost all automated bibliographic systems.[3] Once a document is upended to the library's catalog with proper MARC 21 format, the visibility of the document increases.[4]

The importance of the MARC 21 control field and its effective implementation in university library OPACs in Karnataka were studied by Chandrappa, Narayanaswamy, and Harinarayana.[5] One thousand and eighty-eight (1,088) records, an average of 121, were selected from the nine university library web OPACs for the study using a random method.

Draper and Lederer identified the high-impact problems and solutions derived for multiple 245 (title) fields; titles beginning with a, an, or the without any gap; dollar sign used in the text; fixed field date; languages; subject headings; creating proxy URLs; classification numbers; and author authority control (e.g., corporate entries and presidential entry errors) while improving the data for the library catalog.[4]

It is essential to notice that in most of the literature reviewed, the researchers selected libraries with web OPACs to assess library catalog records. The literature review advocates local-level investigation of catalog records, and there needs to be more such local-level investigation in Assam. Reviewed attempts check the utilization of selected tags and specific number of records. However, there need to be more attempts where all the catalog records are analyzed. Using MarcEdit software for checking the quality of bibliographic records was not traced. This study will fulfill the need for local-level assessment of the catalog records in the colleges of Assam. This study shows the utilization, extent, and consistency of the MARC 21 tags in the catalog records of the college libraries of Assam.

Objectives

Quality is something that describes the level of excellence of service or performance. In India, the National Assessment and Accreditation Council (NAAC) estimates the quality of the library as part of their assessment of the quality of higher education institutions.[6] Quality practices in libraries have a key influence in improving the standard of the parent institute. The quality of a catalog is directly proportional to the retrieval efficiency. Thus, the objectives of this study are:

- To investigate the presence of duplicate records in the catalog.
- Trace the utilization of tags used by the libraries studied, i.e., study their utilization and consistency.
- To study the utilization of the most frequent MARC 21 tags of the selected college libraries as identified by the Library of Congress.
- To validate the catalog through MarcEdit and analyze the results.

Materials and methods

Catalog records of six libraries of those colleges affiliated with the State Universities of Assam and accredited by NAAC with a minimum B grade are identified. While collecting the data, only those catalogs with MARC-compatible records are selected and captured through the web-OPAC of the library. Collected files are in .mrc format and later converted to .csv format using MarcEdit for analysis. MarcEdit and Microsoft Office Excel, with different add-in provisions like "Fuzzy lookup" etc., are used for general and primary records investigation. The present survey tries to understand the catalog information required by the local users.

Data from the following libraries in Table 1 are scraped for the current study.

Analysis

Duplicate checks and unique records in the library catalogs

A duplicate catalog record is considered the complete carbon copy of a record that already exists and is added again as a separate record. Duplicate records make the catalog data dirty.[7] This kind of data reduces efficiency in proportion to the actual number of records in the database. Duplicate records affect decision-making in libraries. It also has an impact on library automation cost-effectiveness through indirect expenditures of time, i.e., searches and redundant cataloging and direct expenditures of fees, i.e., outsourcing, data migration, searches, interlibrary loan, etc.[8]

Table 1. Details of the data collected from the colleges.

Name of the College	Address	Year of establishment/ Affiliated with	NAAC Accreditation with	Name of the Library	Total cataloged records	Data Collected on	Code for the study
Jagannath Barooah College (Autonomous)	Jorhat, Assam	1930/Dibrugarh University	"A" Grade	Krishna Kanta Handiqui Library	36,406	03/12/22	C-1
Arya Vidyapeeth College (Autonomous)	Guwahati, Assam	1958/Gauhati University	"A" Grade	Central Library	22,470	09/01/23	C-2
B. H. College	Barpeta, Assam	1966/Gauhati University	"A" Grade	B H College Library	16,682	29/12/22	C-3
Pandu College	Guwahati, Assam	1962/Gauhati University	"B" Grade	Pandu College Library	16,918	12/01/23	C-4
SPP College, Namti	Sivasagar, Assam	1970/Dibrugarh University	"B" Grade	SPP College Library	12,255	11/11/22	C-5
Bahona College	Jorhat, Assam	1966/Dibrugarh University	"A+" Grade	Bahona College Central Library	12,619	12/12/22	C-6

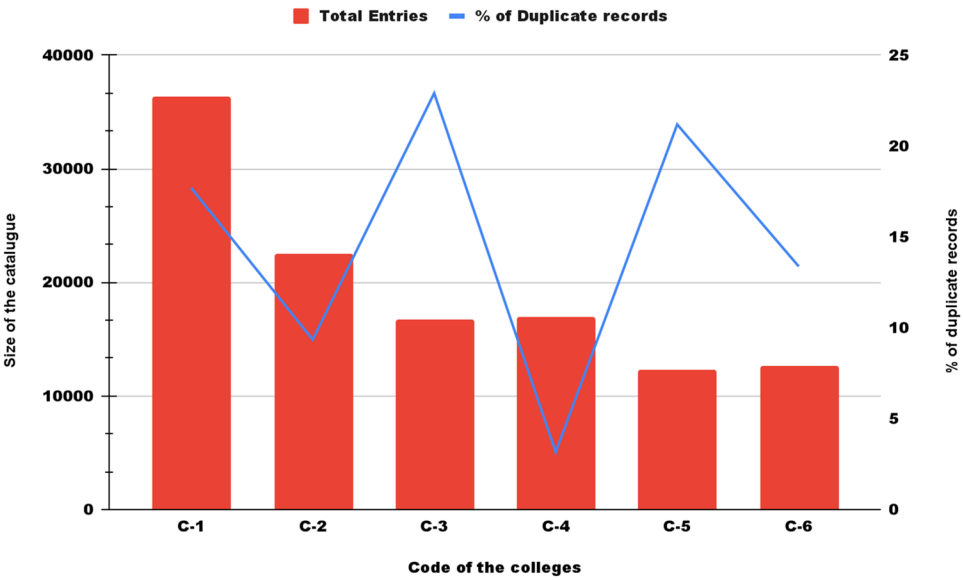

Figure 1. Size of the catalog does not influence the % of duplicate records.

In this study, the sample catalogs have entries ranging from 12,255 to 36,405 and have various percentages of duplicate records. Duplicate records in the sample data range from 3.16% to 22.91%.

Figure 1 shows that the size of the catalog records did not impact the percentage of duplicate records. The awareness and carefulness of the catalogers can reduce the percentage of duplicate records in the catalog.

The following factors influenced the process of identifying duplicates in the catalogs:

1. Typographical dissimilarity for cataloging the same item.
2. Inappropriate use of tags in the catalogs.
3. Incomplete catalogs or omitted information in the previous record.

Utilization of MARC 21 tags used by the libraries

MARC, which aims to provide universally acceptable bibliographic records in machine-readable form, has been the most popular communication format. A MARC record consists of three parts: the leader, which contains coded values based on their position and which defines the processing of the record; the directory, which contains the tags, starting location, and length of each field within the record; and the variable fields. The MARC 21 formats are defined for five types of data, i.e., bibliographic, holdings, authority, classification, and community information.[9]

The basic division of tags for MARC 21 bibliographic records are:

0XX Control information, numbers, codes
1XX Main entry
2XX Titles, edition, imprint
3XX Physical description etc.
4XX Series statements
5XX Notes
6XX Subject added entries
7XX Added entries other than subjects
8XX Series added entries
9XX for locally defined uses such as local barcode numbers, location, document category, etc.

The primary investigation of the tags used by the libraries studied shows that the number of tags used by the libraries ranges from 21 to 67. The highest number of tags utilized by the Library is C-2; however, out of these 67 tags, only five are available in 100% of the records, and altogether 13 tags are available in more than 60% of the data.

MARC 21 tags utilized by all the libraries in cataloging their books are 000(LDR), 001, 005, 008, 020, 082, 100, 110, 245, 250, 260, 300, 650, 942, 952, and 999. In general, the college libraries of Assam utilize eight MARC 21 tags out of the 26 most frequent tags identified by the Library of Congress, i.e., 020, 100, 245, 250, 260, 300, 490, and 650.

From Figure 2, the utilization of primary divisions of MARC 21 differs from library to library. Some libraries attempted to utilize at least a tag from every division, and few were confined to utilizing a minimal number of divisions in their records. Again, within the divisions, the utilization of tags differs. From Figure 3, it is clear that tags from the division 'control information, numbers, and codes' occupy the highest share in the utilization of fundamental division of MARC 21 tags in the bibliographic records, followed by local use, main entry and title, edition and imprints and it is lowest for series statement, added entries other than subjects, and series added entries.

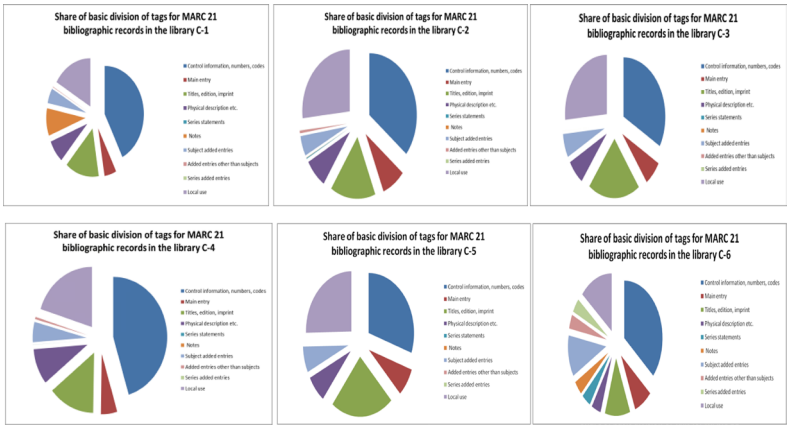

Figure 2. Utilization of tags from basic divisions of MARC 21 in different libraries (C-1 to C-6) in their catalog records.

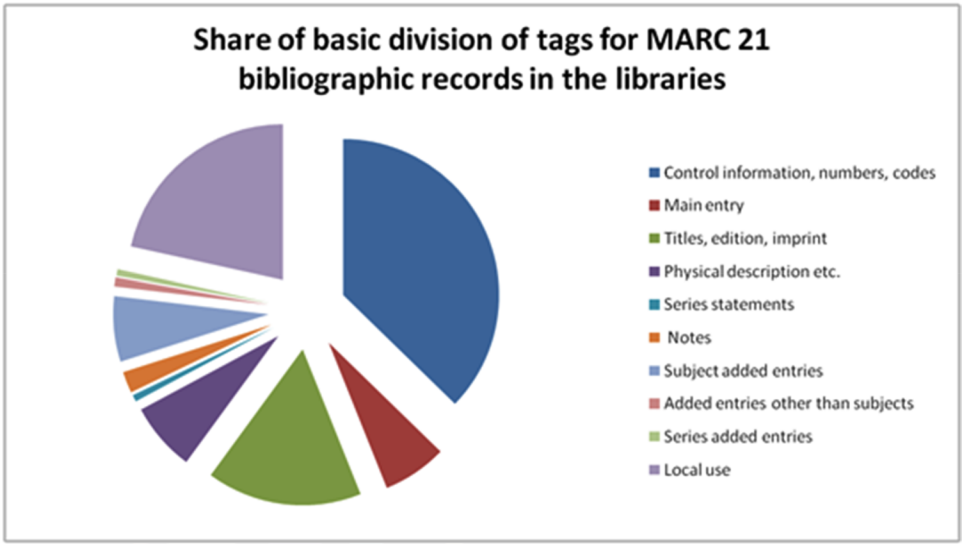

Figure 3. Overall utilization of tags from basic divisions of MARC 21 in different libraries in their bibliographic records.

Consistency in using tags

The extensively used tags are related to control information and local data elements. Control field tags 000, 001, 003, 004, 005, 006, 007, and 008 are system-generated content, and 9XX tags for local data elements. Except for the control fields and local data elements, the numbers of tags used by the libraries C-1 to C-6 are 29, 54, 15, 28, 10, and 18, respectively.

The overall quality of a catalog depends on a combination of factors, i.e., accuracy, consistency, timeliness, and functionality.[10] Table 2 shows that among these MARC 21 tags, only one college has more than 85%

Table 2. List of tags used in more than 90% of the records of the catalog and the immediate next tag to them.

C-1		C-2		C-3		C-4		C-5		C-6	
Tags	% of utilization	Tags	% of utilization	Tags	% of utilization	Tags	% of utilization	Tags	% of utilization	Tags	% of utilization
245	100	245	100	245	99.99	653	111.04	040	100	300	100
650	100	300	99.96	260	99.37	245	100	245	100	653	100
541	100	260	99.74	100	99.08	041	99.72	100	98.41	700	100
082	99.57	0 082	99.74	082	98.92	100	99.71	250	96.32	852	100
300	99.67	100	92.43	300	98.74	082	98.26	260	95.60	245	99.52
260	98.74	**650**	**68.25**	250	95.46	365	96.53	300	93.81	500	99.52
041	98.63			650	94.02	300	96.48	082	93.03	082	96.99
100	98.62			**041**	**60.47**	260	96.12	650	90.18	110	96.99
020	94.73					**250**	**88.37**	**020**	**0.01**	100	96.64
250	**83.75**									490	96.64
										024	96.08
										040	96.08
										041	96.08
										600	96.08
										650	96.08
										260	94.00
										250	**77.50**

consistency in terms of using their utilized tags in the catalog. The consistency rate for the other colleges is 31.03%, 9.25%, 46.66%, 28.57%, and 80%.

Consistency is among the four key elements of the quality of a library catalog and in this paper, we are describing the consistency based on the utilized field tags only because all records may not need to use all of the subfields within tags. Moreover, there is no hard and fast rule to use a minimum or maximum number of MARC 21 tags in a catalog, so the calculation of consistency is calculated based on the field tags which are chosen by the cataloger to be utilized in their catalog, i.e., the consistency of those tags is calculated based on those considered essential by the cataloger and this is different for each of the libraries. For example, C-2 library is using fifty-four field tags excepting the control fields and local data element fields: 010, 015, 016, 020, 022, 024, 025, 035, 040, 041, 042, 050, 072, 082, 100, 110, 111, 210, 240, 245, 246, 250, 260, 300, 350, 440, 490, 500, 501, 504, 505, 506, 510, 520, 521, 533, 538, 580, 583, 600, 610, 650, 651, 655, 700, 710, 740, 752, 773, 776, 800, 810, 830, and 856. From analysis, it is revealed that out of these fifty-four tags only 245, 300, 260, 082, and 100 are used consistently. Field tag 650 is used in 68.25% of the total records whereas it could be utilized in 100% of the documents; this definitely affects document retrieval. Thus, the consistency rate for the college C-2 is 9.25% only.

Extensively used tags

Earlier research revealed that the library web OPACs studied had given scant attention to control fields; 001 and 003 were the two tag numbers used extensively.[5] University libraries must pay more attention to the

control field data elements in their OPACs. Our analysis for all colleges reveals that the libraries extensively use control fields 000, 001, 005, and 008 and tags for local data elements 942, 952, and 999. A total of 117,361 catalog records was analyzed: tags 245, 300, 082, 260, 100, 650, 040, 250, and 041 are used in more than 60% of the records, i.e., 99.94, 98.46, 98.36, 97.80, 97.42, 77.46, 73.53, 70.98, and 63.92%, respectively. Tag 020 is present in 65,076 records (55.44% of the total records) but interestingly, out of these 65,076, 49,093 (75.43%) are contributed by only two colleges (C-1 and C-4). Tag 020 is utilized by C-1 and C-2 in 94.73% and 86.38% of their total records, respectively. This means that though the percent of utilization of tag 020 is more than 50%, it is not uniform in all the colleges. In some colleges, the tag for ISBN is used extensively, whereas some colleges give scant attention to utilizing this tag.

Use of subfield in the extensively used tags

The statistics from Table 3 revealed that the rate of use of these extensively used tags is different among all the colleges. There is diversity in the share percentage for utilization of a particular tag among the colleges. For example, all the most extensively used tags are used rigorously by C-1, whereas C-2 and C-4 do not use tags 250 and 650 rigorously. All the colleges have a very significant percentage of utilization for the tag 040 except C-3. For C-3, the use of tag 040 is nil. Subfields are also used differently by the colleges. Within the same tag, the use of subfields is also different. For example, for tag 650, C-2 uses the subfield codes $a, $y, $z, $x, $2, and $v, whereas C-3, C-5, and C-6 use only one subfield code, i.e., $a. Among the six subfield codes used by C-2 for the tag 650, only $a is extensively used; though the other five subfield codes are there, the number present is fewer than 100 out of 15,336.

Tag 245 Title statement:

The college libraries that are under our study used the following subfields for tag 245:

$a Title (NR)
$b Remainder of title (NR)
$c Statement of responsibility,
$e-
$h Medium (NR)

Tag 245 is the most extensively used tag by the six colleges studied. Among the thirteen subfield codes for this tag, college libraries utilize five. The percentage of utilization of these subfields is different in the records.

Although subfield $e is not in the MARC 21 format as a subfield code, in our analysis, colleges C-1 and C-2 use it. C-1 uses the subfield $e in

those records where $a subfield is filled by vernacular language (text font), i.e., $e is used to enter the title using English text font. On the other hand, C-2 is using the $e code in some records to refer to the author, editor, etc.

Tag 300 Physical Description:

The college libraries that are under our study used the following subfields for tag 300:

$a Extent (R)
$b Other physical details (NR)
$c Dimensions (R)
$e Accompanying material (NR)
$f Type of unit (R)

Among the ten subfield codes for this tag, it is observed that five codes are utilized by the college libraries. The percentage of utilization of these subfields is different in the records. Subfield $a is used in most of the records followed by $c, $b, $e, and $f. Subfield $f is used only twice by C-2 with a value of 40,446. However, according to the MARC 21 format, $f is used for a type of unit only.

Tag 082 Dewey Decimal Classification:

The college libraries that are under our study used the following subfields for tag 082:

$a Classification number (R)
$b Item number (NR)
$2 Edition information (NR)

The cataloger assigns the class number to the document. Class and item numbers together form the call number, which helps the users find their required document, i.e., in information retrieval. Class number is available for 113,760 records in the records we analyzed, whereas item number is available for only 79,623 records. Apart from these two subfield codes, $2 was found in a small dataset.

Tag 260 Publication, Distribution, etc. (Imprint):

The college libraries that are under our study used the following subfields for tag 260:

$a Place of publication, distribution, etc.
$b Name of publisher, distributor, etc.
$c Date of publication, distribution, etc.
$#
${

Along with the subfield codes $a, $b, and $c two inconsistencies, i.e., $# and $ {are traced in the records of C-4 under the tag 260.

Table 3. A statistical representation of subfields in the extensively used tags by the college

Tag 245	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		245	36,406		245	22,470		245
		$b	397		$b	4,201		$b
		$a	36,404		$a	22,470		$h
		$e	572		$e	17		$a
		$c	34,918		$c	14,729		$c
		$h	45		$h	4,619		
Tag 300	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		300	36,181		300	22,462		300
		$b	563		$b	77		$c
		$c	29,797		$a	22,456		$e
		$a	36,136		$f	2		$a
					$e	10		$b
					$c	22,392		
Tag 082	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		082	36,251		082	22,412		082
		$b	36,214		$b	22,349		$b
		$2	4		$2	7,757		$2
		$a	36,253		$a	22,388		$a
Tag 260	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		260	35,948		260	22,412		260
		$b	35,955		$b	22,376		$b
		$c	35,684		$c	20,923		$c
		$a	35,976		$a	22,261		$a
Tag 100	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		100	35,905		100	20,770		100
		$e	30,670		$q	38		$e
		$a	35,905		$9	15,947		$a
		$d	4		$a	20,745		$d
					$d	76		
					$e	12		
Tag 650	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		650	36,709		650	15,336		650
		$z	1		$a	15,324		$a
		$x	3		$y	18		
		$a	36,725		$z	21		
					$x	65		
					$2	2		
					$v	20		
Tag 040	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		040	36,406		040	15,297		040
		$c	3		$c	36		
		$a	36,403		$b	2		
					$a	15,292		
					$d	20		
Tag 250	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		250	30,491		250	348		250
		$b	25,515		$b	1		$b
		$a	30,482		$a	348		$a
Tag 041	C-1			C-2			C-3	
		Tag/Subfield	Total		Tag/Subfield	Total		Tag/Subfield
		041	35,910		041	19		041
		$h	28,953		$a	1		$a
		$a	35,904		$d	18		
		$g	22					

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
16,682	245	16,918	245	12,255		245	12,568	
887	$c	14,577	$c	12,102		$b	1,060	
3,193	$h	9	$h	355		$h	11,839	
16,682	$b	310	$b	35		$a	12552	
1	$a	16,918	$a	12,254		$c	10,720	

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
16,473	300	16,323	300	11,496		300	11,927	
13,805	$c	12,857	$a	11,496		$c	200	
1,400	$b	7,916				$b	200	
16,461	$a	16,312				$a	11,927	
21	$e	1,020						

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
16,504	082	16,625	082	11,401		082	12,249	
4,090	$b	16,516	$b	1		$b	453	
2	$2	1	$a	11,402		$a	12,237	
16,375	$a	15,105						

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
16,578	260	16,263	260	11,716		260	11,872	
16,543	$b	16,043	$b	11,718		$b	11,858	
13,587	$a	16,211	$a	11,720		$c	901	
16,576	$#	3				$	1	
	$c	15,981				$a	10,656	
	${	1						

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
16,530	100	16,870	100	12,060		100	12,205	
12,173	$e	6,610	$e	3		$d	496	
16,571	$a	16,869	$a	12,058		$e	1,404	
181	$#	1	$d	1		$a	12,204	
						$q	155	

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
15,686	650	2	650	11,052		650	435	
15,686	$z	2	$a	11,052		$a	435	
	$a	2						

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
NIL	040	10,214	040	12,255		040	12,134	
	$b	10,214	$a	12,255		$c	12,134	
	$c	10,213				$a	12,134	
	$a	10,213						
	$d	10,213						

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
15,925	250	14,952	250	11,804		250	9,787	
13,778	$b	9,243	$b	3		$b	975	
15,907	$a	14,952	$a	11,801		$a	9,597	

C-4			C-5			C-6		
Total	Tag/Subfield	Total	Tag/Subfield	Total		Tag/Subfield	Total	
10,089	041	16,872	041	NIL		041	1,258	
10,089	$h	136				$a	1,258	
	$a	16,872						

Tag 100 Main Entry-Personal Name:

The college libraries that are under our study used the following subfields for tag 100:

$a Personal name (NR)
$d Dates associated with a name (NR)
$e Relator term (R)
$q Fuller form of name (NR)
$9
$#

Along with the subfield codes $a, $d, $e, and $q two inconsistencies, i.e., $# and $9 are traced in the records. Subfield $9 is used in 15,947 records of C-2 under the tag 100.

Tag 650 Subject Added Entry-Topical Term (R):

The college libraries that are under our study used the following subfields for tag 650:

$a Topical term or geographic name entry element (NR)
$v Form subdivision (R)
$x General subdivision (R)
$y Chronological subdivision (R)
$z Geographic subdivision (R)
$2 Source of heading or term (NR)

Though subfield codes from all three divisions are utilized, only subfield code $a is used in maximum records. C-2 attempted to maintain a detailed record for tag 650.

Tag 040 Cataloging Source (NR):

The college libraries that are under our study used the following subfields for tag 040:

$a Original cataloging agency (NR)
$b Language of cataloging (NR)
$c Transcribing agency (NR)
$d Modifying agency (R)

Though subfield codes $a, $b, $c, and $d are utilized, only subfield code $a is mostly used by the colleges except C-4. C-4 maintained a detailed record for 10,214 records under the tag 040.

Tag 250 Edition Statement (R):

The college libraries that are under our study used the following subfields for tag 040:

$a Edition statement (NR)
$b Remainder of edition statement (NR)

Tag 041 Language Code (R):

The college libraries that are under our study used the following subfields for tag 040:

$a Language code of text/sound track or separate title
$d Language code of sung or spoken text
$g Language code of accompanying
$h Language code of original

Though subfield codes $a, $b, $g, and $h are utilized, only subfield code $a is mostly used by the colleges. C-5 did not utilize this tag.

Indicators

The use of indicators is very poor. College libraries rarely attempted to utilize the indicators. When we analyze the catalogs through the MarcEdit validator we see that most of the available indicators were also not correct.

Requirement of the local users

We have surveyed 150 users of college libraries. It is revealed from the survey that users usually use some bibliographic information to search for their required documents in the library. The following are those bibliographic criteria in descending order of the preference of the users:

Title
Author
Chapter name, i.e., summary note
Subject
Subtitle
Publication details

Moreover, it is also found in the survey that library users use the OPAC to know the availability and the location of the books. Usually, they do not search for the classification number. But to give the document a systematic and distinct location on the shelf the cataloger needs to classify the document properly. Fourteen and six-tenths percent of the respondents often and 12% of the respondents sometimes utilize the ISBN to search for books in the library. The remaining 73.4% never utilize ISBN to search for a book in the library.

Results

The following conclusions were drawn from the examination of the data gathered from six colleges from two different universities with three different grades:

i. Duplicate records can adversely affect the catalog. The presence of duplicate records, up to 22.91%, is traced in the college libraries of Assam. The

librarians should be careful in this regard and supervise the catalogers while cataloging to make the catalog duplicate accessible.

ii. The number of MARC 21 tags used by the libraries ranges from 21 to 67. MARC 21 tags 000 (LDR), 001, 005, 008, 020, 082, 100, 110, 245, 250, 260, 300, 650, 942, 952, and 999 are utilized by all the libraries. However, only one college has more than 85% consistency in using their utilized tags in the catalog. The college libraries should be more consistent in using MARC 21 tags.

iii. Tags 245, 300, 082, 260, 100, 650, 040, 250, and 041 are identified as extensively used tags. However, our survey found that notes and ISBN are two crucial components for information retrieval from the lens of the users. Thus to satisfy the potential users of the libraries, the college libraries of Assam should utilize tags from the 5XX group and 020 tags extensively.

iv. A total of 117,361 catalog records were analyzed and it was found that out of the 26 most frequent MARC 21 tags identified by the Library of Congress, the college libraries are using only eight, i.e., 020, 100, 245, 250, 260, 300, 490, and 650.

v. The use of indicators and subfield tags is inconsistent and errors were found during MarcEdit validation in the college libraries of Assam. The librarian should be more cautious in these two aspects.

Discussion

MARC 21 is very popular and widely used in libraries across the globe. MARC 21 is an international standard for bibliographic description, and enables interoperability, data manipulation, copy cataloging, etc., among libraries and information service platforms. Most of the OPACs used by the academic libraries of India offer the primary search features users require.[11] A quality catalog with accurate MARC 21 records can enhance information retrieval from the library's catalog. The catalogs can acquire optimum quality through the extensive use of MARC 21 tags with consistency and accuracy. The use of fewer MARC 21 tags means less complete bibliographic detail, which affects information retrieval.

With the advent of time, cataloging has become a costly practice. Duplicate records in a catalog affect decision-making and cost-effectiveness through different expenditures. In our study, duplicate records are traced, and the libraries must address this issue. At least those MARC 21 tags that can fulfill the needs of the local users must be used by the libraries. Consistency and accuracy are two important parameters of a quality catalog. The libraries should use a standard number of MARC 21 tags in their catalogs with consistency and accuracy to maintain a quality catalog that meets international standards. If college libraries can do so, they can open up their resources for copy catalogs. In such a scenario, information

retrieval will be better, and cooperation among the libraries will increase, ultimately saving the libraries' time, space, and money.

Assam's college libraries use MARC 21 bibliographic records in their catalogs, which are accessible through an OPAC and a web-OPAC. It was also observed during our study that colleges are gradually adopting web-OPACs in their libraries. Evaluation of the current activities is a tool for continuous improvement. In this study, we have identified the areas where major and frequent errors are made by the catalogers and prevention can be emphasized or a scope for improvement is identified. This study fulfilled the need for local-level assessment of the catalog records. Rule-based validation must be practiced regularly to develop and maintain standard catalogs in Assam's college libraries. This study can be considered a local assessment of the catalogs of college libraries in Assam.

Conclusion

Our paper reveals that though the college libraries are partially fulfilling the requirements of their potential users, the utilization of MARC 21 bibliographic standards in terms of extent and consistency has scope to improve. Users felt it would be better if the catalogs had more information on a particular document, like indexing, proper class number, and ISBN, among others. This study brings light to the existing practice of cataloging in the college libraries of Assam in terms of the utilization of the MARC 21 standard and its structure, i.e., the tags, subfield codes, and indicators. Places where they should be more careful and which divisions could be more helpful for them are identified and suggested in the paper. In India, libraries have a significant role in obtaining better accreditation for their institute in NAAC assessment.[12] But it is revealed from the study that the NAAC grade is not proportional to the quality of the catalog. The knowledge and effort of the librarian and cataloger influence the utilization of the MARC 21 tags in their catalogs. This study has a future scope of studying the accuracy of the catalog contents to complete the local-level assessment of college catalogs in Assam.

Acknowledgment

The authors acknowledge the respondents in the survey for their consent, time, and valuable input during the survey.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Statement of declaration

It is to certify that the authors have no affiliations with, or involvement in, any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript. The authors received no financial support for the research, authorship, and/or publication of this article. The article is the authors' original work and hasn't received prior publication.

ORCID

Bidyut Bikash Boruah http://orcid.org/0000-0002-5773-7658
S. Ravikumar http://orcid.org/0000-0002-2317-1596
Fullstar Lamin Gayang http://orcid.org/0000-0002-0980-2443

Notes

1. K. M. Chollom and Daniel Abubakar, "Importance of Cataloguing and Classification in Libraries," *International Journal of Research Multi-Disciplinary Studies* 1, no. 1 (2013): 108.
2. *Understanding MARC: Bibliographic Machine-Readable Cataloging. Part VII: A Summary of Commonly Used MARC 21 Fields*, 8th ed. (Washington, DC: Network Development and MARC Standards Office, Library of Congress, 2009), accessed June 8, 2022, https://www.loc.gov/marc/umb/um07to10.html.
3. Michael Gorman, "Cataloguing in an Electronic Age," *Cataloging & Classification Quarterly* 36, no. 3–4 (2003): 5–17, https://doi.org/10.1300/J104v36n03_02.
4. Daniel Draper and Naomi Lederer, "Analysis of Readex's Serial Set MARC Records: Improving the Data for the Library Catalog," *Government Information Quarterly* 30, no. 1 (2013): 87–98, https://doi.org/10.1016/j.giq.2012.06.010.
5. Chandrappa, Narayanaswamy Vasantha Raju, and N. S. Harinarayana, "Use of MARC21 Control Field Data in University Library OPACs in Karnataka: A Study," *Annals of Library and Information Studies* 68 (2021): 21–7.
6. Anita Chandwani, "Role of Libraries and Librarians in NAAC Assessment," *Challenges and Future Prospects in Online Submission of SSR* 108 (2022), http://eprints.rclis.org/44074/1/DAYANAND%20ARTICLE.pdf.
7. Nicholas Samuel, "10 Reasons How Duplicate Data Harms Your Business," HEVO (April 5, 2022), https://hevodata.com/learn/duplicate-data.
8. E. T. O'Neill, S. A. Rogers, and W. M. Oskins, "Characteristics of Duplicate Records in OCLC's Online Union Catalog," *Library Resources & Technical Services* 37, no. 3 (1993): 59–71.
9. D. Das, "MARC 21: The Standard Exchange Format for the 21st Century," *2nd Convention PLANNER – 2004* (Ahmedabad: INFLIBNET Center, 2004), 154–63.
10. Ann Chapman and Owen Massey, "A Catalogue Quality Audit Tool," *Library and Information Research News* 26, no. 82 (2002): 26–37.
11. Kanta Kapoor and O. P. Goyal, "Web-Based OPACs in Indian Academic Libraries: A Functional Comparison," *Program: Electronic Library and Information Systems* 41, no. 3 (2007): 291–309, http://doi: 10.1108/00330330710774165.
12. Jange Suresh, "NAAC Accreditation and Academic Libraries: Librarians' Role Beyond Librarianship," *Journal of Indian Library Association* 57, no. 4 (2021): 12.

Appendix

Subfield codes as prescribed in MARC21 for the tags that are extensively used by the libraries.

Subfield codes used for the tag 245 as mentioned by the Library of Congress are:

$a—Title (NR)
$b—Remainder of title (NR)
$c—Statement of responsibility, etc. (NR)
$f—Inclusive dates (NR)
$g—Bulk dates (NR)
$h—Medium (NR)
$k—Form (R)
$n—Number of part/section of a work (R)
$p—Name of part/section of a work (R)
$s—Version (NR)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)

Subfield codes used for the tag 300 as mentioned by the Library of Congress are:

$a—Extent (R)
$b—Other physical details (NR)
$c—Dimensions (R)
$e—Accompanying material (NR)
$f—Type of unit (R)
$g—Size of unit (R)
$3—Materials specified (NR)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)

Subfield codes used for the tag 082 as mentioned by the Library of Congress are:

$a—Classification number (R)
$b—Item number (NR)
$m—Standard or optional designation (NR)
$q—Assigning agency (NR)
$2—Edition information (NR)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)

Subfield codes used for the tag 260 as mentioned by the Library of Congress are:

$a—Place of publication, distribution, etc. (R)
$b—Name of publisher, distributor, etc. (R)
$c—Date of publication, distribution, etc. (R)
$e—Place of manufacture (R)
$f—Manufacturer
$g—Date of manufacture (R)
$3—Materials specified (NR)
$6—Linkage (NR)
$8—Field link and sequence number (R)

Subfield codes used for the tag 100 as mentioned by the Library of Congress are:

$a—Personal name (NR)
$b—Numeration (NR)
$c—Titles and words associated with a name (R)
$d—Dates associated with a name (NR)
$e—Relator term (R)
$f—Date of a work (NR)
$g—Miscellaneous information (R)
$j—Attribution qualifier (R)
$k—Form subheading (R)
$l—Language of a work (NR)
$n—Number of part/section of a work (R)
$p—Name of part/section of a work (R)
$q—Fuller form of name (NR)
$t—Title of a work (NR)
$u—Affiliation (NR)
$0—Authority record control number or standard number (R)
$1—Real World Object URI (R)
$2—Source of heading or term (NR)
$4—Relationship (R)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)

Subfield codes used for the tag 650 as mentioned by the Library of Congress are:

Main term portion
$a—Topical term or geographic name entry element (NR)
$b—Topical term following geographic name entry element (NR)
$c—Location of event (NR)
$d—Active dates (NR)
$e—Relator term (R)
$g—Miscellaneous information (R)
$4—Relationship (R)

Subject subdivision portion
$v—Form subdivision (R)
$x—General subdivision (R)
$y—Chronological subdivision (R)
$z—Geographic subdivision (R)

Control subfields
$0—Authority record control number or standard number (R)
$1—Real World Object URI (R)
$2—Source of heading or term (NR)
$3—Materials specified (NR)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)

(*Continued*)

Appendix Continued

Subfield codes used for the tag 040 as mentioned by the Library of Congress are:

$a—Original cataloging agency (NR)
$b—Language of cataloging (NR)
$c—Transcribing agency (NR)
$d—Modifying agency (R)
$e—Description conventions (R)
$6—Linkage (NR)
$8—Field link and sequence number (R)

Subfield codes used for the tag 250 as mentioned by the Library of Congress are:

$a—Edition statement (NR)
$b—Remainder of edition statement (NR)
$3—Materials specified (NR)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)

Subfield codes used for the tag 250 as mentioned by the Library of Congress are:

$a—Language code of text/sound track or separate title (R)
$b—Language code of summary or abstract (R)
$d—Language code of sung or spoken text (R)
$e—Language code of librettos (R)
$f—Language code of table of contents (R)
$g—Language code of accompanying material other than librettos and transcripts (R)
$h—Language code of original (R)
$i—Language code of intertitles (R)
$j—Language code of subtitles (R)
$k—Language code of intermediate translations (R)
$m—Language code of original accompanying materials other than librettos (R)
$n—Language code of original libretto (R)
$p—Language code of captions (R)
$q—Language code of accessible audio (R)
$r—Language code of accessible visual language (non-textual) (R)
$t—Language code of accompanying transcripts for audiovisual materials (R)
$2—Source of code (NR)
$6—Linkage (NR)
$7—Data provenance (R)
$8—Field link and sequence number (R)