DESIDOC Journal of Library & Information Technology, Vol. 36, No. 6, November 2016, pp.431-436 DOI: 10.14429/djlit.36.6.10608

Investigation into Diligence in Metadata Records of Mysore University Library

Chandrappa* and N.S. Harinarayana**

Department of Studies in Library and Information Science, University of Mysore, Karnataka - 570 005
E-mail: chandruekm@gmail.com; ns.harinarayana@gmail.com

ABSTRACT

Accuracy in cataloguing records is of paramount importance in catalogues. It is not difficult to imagine that errors in the metadata hamper the retrieval of bibliographic records. The present study attempts to investigate the error rate found in the metadata records of the Mysore University Library (MUL). The MUL is one of the oldest university libraries in the country, India. The study selected and analysed the accuracy of 624 metadata records of the Mysore University Library along with 491 records of Library of Congress (LoC). The assessment of the selected records discerned two main types of errors notably: Major and minor errors. These findings illustrate and prove that the percentage of error rate in MUL is higher than that of LoC. Consequently, the study highlighted that the prevailing error ratio between both institutions reaches almost up to 5:1 (MUL: LoC).

Keywords: Metadata records analysis, metadata quality, metadata evaluation, metadata records investigation, metadata creation, OPAC quality, Mysore University Library, Library of Congress

1. INTRODUCTION

Accurate metadata records are the backbone of user services in libraries. A purview of the history in the field underpins that there have been some efforts to enhance the quality of metadata records regarding accuracy and completeness. The invention of a myriad of cataloguing codes is one such enterprise to ensure the quality, standard, and consistency in the metadata records. Accurate and comprehensive catalogue data bear a direct influence on the effectiveness of user search of the metadata inacunate inconstent. The presence of errors in 'records/data' dwindle the efficiency of the retrieval capability of the catalogue. This study concentrates on studying the accuracy of metadata in the bibliographic records of Mysore University Library (MUL). The corresponding records in LoC were tested for accuracy and comprehensiveness.

The study analyses the metadata records for the errors such as incorrect bibliographic information, omitted elements, misspelling, MARC code errors and deviations from cataloguing rules. An inaccurate or inconsistent metadata compels readers to perform repeated searches either to find the records they seek or to find everything of interest. The inaccuracies in metadata cast doubt on the quality of other accessible records and offer a bad impression of the library's services.

2. LITERATURE REVIEW

Thomas[1] defines quality as encompassing consistency, depth, appropriateness and timeliness as well as accuracy. Even if perfection is not desirable, and users are satisfied with only partial success in retrieval, high standards and accuracy of library catalogues are essential. Libraries themselves need accurate inventories of their collection for providing the services. OCLC records have been the benchmark for studying the accuracy of metadata. Zeng[2-3] has based his examination on a randomly selected set of 1306 records related to the Chinese language records. The study considers the OCLC sample which composed of 853 records OCLC CJK and 453 RLIN CJK member-contributed records. Zeng[2-3] has identified the common errors and systematized them into three classes: format, content, and editing and inputting errors. Shin[4] study examined only monographic records through the randomly selected from 623 records in which OCLC sample had 508 records, and the remaining 115 records were from RLIN. Fung-yin[5] conducted a similar study in which the researcher analysed 380 Chinese monographic records selected from OCLC's WorldCat (the OCLC Online Union Catalogue) which were processed between October 1995 to February 1996 in the Asian Library of the University of Illinois at Urban-Champaign. Jia[6] made an effort to identify the errors in the metadata. They have examined the randomly selected set of 454 OCLC metadata records. Errors were found in nearly half of the records in the sample.

Benchmarks other than OCLC have been used in some studies. Manaf & Rahman[7] studied the Malaysian National Library CIP comparing with the National Library OPAC by using the discovery method. The data indicates that their CIP bibliographic records are different from those from the National Library OPAC in some cases such as main entry. The study extracts a total of 341 CIP records (January to April 2000) in an attempt to observe, compare and evaluate the CIP records. Zarei[8]

Received 3 September 2016, revised 26 October 2016, online published 30 November 2016

reviews the consistency of information elements status of 163 CIP records and compression with Catalogue after Publication (OPAC) records of the Iranian National Library Bibliographic Center in between 1997 to 2007.

The cataloguing standards, such as AACR and ISBD, have been the reference points for checking the accuracy of bibliographic description. Ajis & Manaf[9] chose a set of 500 samples of the bibliographic records which were analysed based on their compliance with the International Standards like AACR2 and MARC 21. Another similar study was conducted by Enang[10] whereby the National bibliographic agencies create records for 250 publishers issued in each country, using ISBD as a guide.

The sample size is one of the issues that drew the attention of the researchers. Massey & O'Brien[11] and Chapman & Massey[12] have argued that the number of records in the sample depends on the acceptable margin of error in the result, but it does not depend on the size of the population (provided fewer than 10% of records are to be sampled, which is almost always the case). Hewitt[13] has taken 2500 records for his study. Mansor[14] has investigated the variant metadata practices that affect the development of a cooperative cataloguing program in Malaysia. 410 MARC records were sampled from the OPAC databases of three university libraries in Malaysia. This study focused on the selected MARC fields, i.e., that is for the name, title and subject access points.

3. METHODOLOGY

The present study was conceived to investigate the quality of MUL metadata records regarding its accuracy and comprehensiveness of resource description. The Mysore University Library (MUL) has around 4.79 million printed books; wherein in this study, a sample of 624 items have been examined. The metadata records were selected using random sampling. The major reason for restricting the sample size is the constraints of time and financial resources available to carry out the study.

A set of error category was developed to facilitate the categorisation of errors occurring in the sample records. The errors have been classified broadly into two groups: Major and minor. In a bibliographic record, this study considers that major errors include: entry omission, misspelling and wrong entry, misplacing entry, hyphen inserted at wrong places and errors in the transliteration of ISBN. Minor errors, on the other hand, consist of missing full stops and punctuation, inadequate spacing, incorrect upper and lower case so on and so forth.

A systematic approach is used for the data collection. The fifth book (as on 15-07-2015 to 19-07-2015) from the left side of the middle compartment of every rack available in the stack area (including the three floors) and the active-stack area formed the bibliographic entity to be examined for the study. Through this method, the researchers ensured that at least one book from each rack of MUL was taken for the study. The metadata of 624 documents as found in the OPACs of MUL and Library of Congress (LoC) were compared for further analysis. MUL OPAC was searched for each item in the sample through the 'Accession Number'–the unique identifier given for each book procured in the library, and LoC OPAC was searched through the title the 624 records, the metadata of 491 (78.69 %) were found MUL OPAC. It was surprising to note that the database does not still contain the records about 21.31 % of items from the sample. On investigating into the reasons for such omissions, it was found out that the retrospective conversion of the library stack is in progress. Hence, it is unlikely that all records are available on OPAC. As we understand from the concerned authorities, the conversion process is completed for the collection on the first floor of the stack area and all the collection of active-stack in other floors. Hence, it may be assumed that there is a drop in the overall percentage of books found in the OPAC. However, it was found that even in the areas where the conversion process is over the retro-conversion is probably not exhaustive yet. For instance, out of the 272 books from the first floor of the stack area, the researchers could get only 236 (86.76 %) records in the MUL OPAC. It is unfair or too early to pass the judgment on the completeness of the coverage as the retro-conversion project is still in progress.

Only those records which were found in MUL were searched in LoC so that comparison is possible. As said earlier, the LoC records are considered as the benchmark for comparison. Thus, only 491 records were considered for searching in LoC catalogue. Special care has been taken to verify that the same edition of the books from both MUL and LoC were compared. The initial data element used for searching LoC was the title of the document. To identify the specific matching records in LoC, the other details such as author, edition, and year of publication among others considered. In all 350 (71.28%) out of 491 records were available in LoC. These records established the basis for further analysis. It is worth stating that 141 books in MUL were not found in LoC. One of the reasons is that from the 141 books, 50 are books published by vernacular publishers. However, it is not surprising that these books were not found in the LoC database. However, the researchers could not get convincing reasons for the remaining 91 foreign authored books (mostly published by publishers from the western countries) which were available in MUL but absent in LoC.

4. FINDINGS AND ANALYSIS

Metadata as found in both libraries–MUL and LOC, were compared for their differences. Appropriate tables were created in Microsoft Excel.

4.1 Accuracy in Data Entry Regarding ISBNs

The ISBN tag is a key element in the bibliographic description of monographs. It is also predominantly used by for retrieval of books and similar items. The study

Table 1. Accuracy rate of ISBN

S. No.	Name	MUL (per cent)	LoC (per cent)
1.	Major error	23 (14.84)	-
2.	Minor error	-	-
3.	No error	132 (85.16)	142 (100)
	Total	155 (100.00)	142 (100)

looked into the accuracy in data entry as far as ISBNs are concerned in Table 1.

From the 491 records considered for the study, 297 (60.49 %) records possessed/bore the ISBN tag in their metadata records. It was found that the records found in the LoC database are error free. On the other hand, 14.84 % of the records found in the MUL metadata consisted of major errors. As the ISBN tag is significant,

Example: Record 1. ISBN field variants

Record no.	MUL record	LoC record
04	0-12-215450-9	0122154509
20	0-521-21515-3	|a 0521215153
294	0 19 212582 6	|a 0192125826

it is imperative that ISBN numbers should be error-free in metadata records. Example- Record 1 shows the ISBN field variants.

4.2 Accuracy Rate of Main Entry

The MARC tag 100 contains the personal name as the main entry, whereas 110 is for corporate name. According to AACR2 rules (even the latest RDA), the main entry is assigned to the person/institute chiefly responsible for the intellectual or artistic content of the resource. The cataloguer pays utmost importance for the preparation of this metadata element. Any error in this element may prove costly in the retrieval of the resource.

It can be observed from Table 2 that 413 out of 491 records have the main entry tags in their metadata records. MUL has more than 50.00 % errors in their main entry data field. The majority of records (98.28 %) examined in the LoC were found to be error free. This comparison of MUL and LoC gives a stern warning to MUL to pay a very serious attention to the creation of metadata records. Example-Record 2 shows the variants of Main entry field.

4.3 Accuracy Rate of Entry Field Title

The titles of the resource identify the works. Authors

Table 2. Accuracy rate of main entry – (personal name, corporate name)

S. No.	Type of error	MUL (per cent)	LoC (per cent)
1.	Major error	30 (7.26)	5 (1.72 %)
2.	Minor error	206 (49.88)	-
3.	No error	177 (42.86)	286 (98.28)
	Total	413 (100.00)	291 (100.00)

Example: Record 2. Main entry field variants

Record no.	MUL record	LoC record
13	Sarsvati, Chennakesavan	Chennakesavan, Sarasvati, |d 1918-
19	Shah, Sirdar Ikbal Ali	|a Shah, Ikbal Ali, |c Sirdar.
32	Rajni Chadha	|a Chadha, Rajni, |c Dr.
127	NCAER	|a National Council of Applied Economic Research.

pay a lot of attention in giving titles to their works. Diligence is warranted in data entry of this metadata element in a record. It was hypothesised that the number of errors in this field would be minimal.

Table 3 depicts the accuracy rate of entry field title statement (245) in LoC and MUL online metadata. It was found that 93.28 % contains one or the other type of errors in the title statement data field in both the libraries together. 44.6 % of the records found in MUL consists major errors in the title statement data field. Almost 50 % of records bear minor errors in the title statement data field. It is evident from Table 3 that almost all the records in MUL consist one or the other type of errors in the title statement. On the other hand, 73.71 % of the records found in LoC are error free in the title statement data field.

Table 3. Accuracy rate of title statement

S. No.	Name	MUL (per cent)	LoC (per cent)
1.	Major error	219 (44.60)	91 (26.00)
2.	Minor error	239 (48.68)	1 (0.29)
3.	No error	33 (6.72)	258 (73.71)
	Total	491(100.00)	350 (100.00)

Example: Record 3. Title statement field variants

Record no.	MUL record	LoC record
05	Conceptions of personality: Theories and research	|a Conceptions of personality; |b theories and research |c [by] Leon H. Levy.
13	Concepts of Indian philosophy	a Concepts of Indian philosophy / |c Sarasvati Chennakesavan.
31	Urban anthropology : Cross-cultural studies of urbanisation	|an Urban anthropology; cross-cultural studies of urbanisation. |c Edited by Aidan Southall.
249	Life In Letters Of William Dean Howells	|a Life in letters of William Dean Howells. |c Edited by Mildred Howells.

The results indicate that there is an urgency to check the whole database for the accuracy of title data in MUL OPAC. The inaccuracies hamper the retrieval of information based on the title. An example of kind of errors found in MUL OPAC regarding. The variants of title statement are shown in Example-Record 2.

Table 4. Accuracy rate of edition statement

S. No.	Name	MUL (per cent)	LoC (per cent)
1.	Major error	25 (51.02)	2 (5.13)
2.	Minor error	-	1 (2.56)
3.	No error	24 (48.98)	36 (92.31)
	Total	49 (100.00)	39 (100.00)

4.4 Accuracy Rate of the Edition Statement

It is not uncommon to see a work appearing in different versions. The edition statement (tag 250 in MARC) contains the information related to various versions of a resource. Edition information is an individualising element for uniquely identifying the resource and hence requires a lot of attention while recording its details in the database. Table 4 illustrates the accuracy rate of the edition statement found in the metadata records of MUL and LoC databases. On one hand, it was found that major errors in edition statement data field were to the tune of 51.02 % in MUL records which is quite high by any standards. On the other hand, LoC has an accuracy rate of 92.31 % about the data field edition statement (250) which is at a satisfactory level.

Example: Record 4. Edition statement field variants

Record no.	MUL record	LoC record
88	Second edition	|a 2nd ed.
174	Second edition	|a 2nd ed.

As stated for other elements, MUL falls behind LoC as far as accuracy of data entry in tag 250 also. The kinds of errors found are shown below as a sample. Examples of edition statement field variants in record 4.

4.5 Accuracy Rate of Publication Description

Table 5 depicts the publication description data (260) as found in MUL and LoC metadata records respectively. The accuracy of data elements such as year of publication, publishers name and place of publication have been scrutinised in the metadata records. It was on one hand found that 28.68 % of the records certain minor errors. On the other hand, LoC records contain very few errors in relations to the publication description data element. Examples of publication description field Examples of publication description field are shown in Example-Record 5.

The bibliographical description is not complete without recording the details related to the collation which includes preliminary and text pages, illustration details, accompanying materials, etc. Though the information contained in this metadata element is not normally used for retrieval of resource, they are certainly useful for comprehensively and uniquely describing a resource. Hence, diligence is expected.

Table 5. Accuracy rate of publication description

S. No.	Name	MUL (per cent)	LoC (per cent)
1.	Major error	23 (4.68)	4 (1.14)
2.	Minor error	131 (28.68)	5 (1.43)
3.	No error	337 (68.64)	341 (97.43)
	Total	491 (100.00)	350 (100.00)

Example: Record 5. Publication description field variants

Record no.	MUL record	LoC record
05	New 11ans : Random House 1970	New York, Random House [1970]
31	NY: OUP, 1973.	|a New York, |b Oxford University Press, |c 1973.

4.6 Accuracy Rate of Physical Description

Table 6 depicts the physical description data as found in MUL and LoC metadata records. Almost 20 % records in MUL database include major errors. However, 10 % of the LoC records certain major error about physical description. A minor error was also found in MUL and LoC metadata records as well. 22.40 % records in MUL and 4.86 % records in LoC possess minor errors in their metadata records, respectively. The physical description is though not a major access point to retrieve documents, but is an important element in cataloguing practice/records which cannot be ignored while describing information resources.

Example: Table 6. Accuracy rate of physical description

S. No.	Name	MUL (per cent)	LoC (per cent)
1.	Major error	98 (19.96)	35 (10.00)
2.	Minor error	110 (22.40)	17 (4.86)
3	No error	283 (57.64)	298 (85.14)
	Total	491(100.00)	350 (100.00)

The difference regarding errors between MUL and LoC are shown below. Examples of physical description field Examples of physical description field variants can be seen in Example-Record 6.

The second and subsequent elements in the statement

Example: Record 6. Physical description field variants record

Record no.	MUL record	LoC record
05	ix, 492p 22cm	ix, 492 p. |b illus. |c 25 cm.
10	xi, 515 p. ; 23 cm.	|a 515 p. |b illus. |c 24 cm.
18	422p. ; 24 cm.	|a 4 v. |c 26 cm.

Table 7. Accuracy rate of added entry-personal name

S. No.	Name	MUL (per cent)	LoC (per cent)
1.	Major error	38 (22.89)	13 (10.92)
2.	Minor error	61 (36.75)	-
3.	No error	67 (40.36)	106 (89.08)
	Total	166 (100.00)	119 (100.00)

of responsibility appear in the MARC tag 700. Like the field 100, this metadata element contains the details of authors other than the first/major author indicated in the resource. It is needless to emphasise that accuracy is paramount in recording the details in this element also.

4.7 Accuracy Rate of Added Entry

Table 7 indicates the quantum of errors found in added entry personal name data field in metadata records of MUL and LoC's online database. It was found that 60.00% of the records in MUL database include one or the other type of errors (major and minor errors) in added entry–personal name data element. LoC, on the other hand, contains significantly lesser error rate as depicted by the data collected for this study. It was found that the error rate in LoC was only 11.00%.

Added entry-personal name is a key access point for document retrieval in libraries; users would approach many times with a second or third or through an editor for retrieving books from the library. Consequently, it is essential to have added entry element error free in metadata records. The type of errors that were found are shown in Example-record 7.

Example: Record 7. Added entry field variants record

Record no.	MUL record	LoC record
04	Dickson W. Patrick	Dickson, W. Patrick

4.8 Total Error of Metadata Records

Table 8 exposes an expressed error rate that MUL records possess the average error rate of 2.58 %, and LoC records hear the average rate of 0.50 % from the study, It was detected that the error ratio between the institutions is almost 5:1 (MUL: LoC).

Table 8. Total error of metadata records

S. No.	Name	Number of records	Number of total errors	Average error per record
1.	MUL	491	1268	2.58
2.	LoC	350	174	0.50

5. CONCLUSIONS

A comparative study of the metadata records was conducted for evaluating the accuracy of the selected bibliographic databases from two agencies. Quality in metadata is a vital prerequisite in library services for both patrons and staff. The general availability of all library resources is directly affected by the quality of the library database.

The purpose of this study was to alert metadata creators about some major problems in metadata records and to offer choose for improvement. Ideally, the suggestions given in the study will assist metadata professionals to avoid these errors altogether. This study can also be beneficial in the training of new metadata creators and may inspire other researchers to investigate the quality of metadata records through other innovative tools.

REFERENCES

1. Thomas, S.E. Quality in bibliographic control. *Library Trends*, 1996, **44**(3), 491-505.
2. Zeng, L. Quality control of Chinese-language records using a rule-based data validation system-part 1: An evaluation of the quality of Chinese-language records in the OCLC OLUC database. *Catalog. & Classif. Quart.,* 1993, **16**(4), 25-66.
3. Zeng, L. Quality control of Chinese-language records using a rule-based data validation system-part 2: A study of a rule-based data validation system for online Chinese cataloging. *Catalog. & Classif. Quart.*, 1994, **18**(1), 3-26.
4. Shin, H. Quality of Korean cataloguing records in shared databases. *Catalog. & Classif. Quart.*, 2003, **36**(1), 55-90.
5. Fung-yin, K.S. Quality control of Chinese monographic records: a case study. *J. of East Asian Lib.,* 1998, **116**(1), 31-40.
6. Jia, B. An empirical study on OCLC catalog record errors for the copy cataloging of Chinese monographs. University of North Carolina, Chapel Hill. 2007. PhD Thesis.
7. Abd Manaf, Z. & Abdul Rahman, R. Examining the quality of national library of Malaysia (NLM) cataloguing in publication (CIP) records. *Library Review,* 2006, **55**(6), 363-73.
8. Zarei, H. Comparison of cataloguing in publication (CIP) with cataloguing after publication (OPAC) of Iran national library bibliographic center. *Lib. Phil. and Pract.,* 2012, 794.
9. Md Ajis, A.F. & Abd Manaf, Z. Appraising the quality of Malaysian union catalog. *Lib. Phil. and Pract.* (e-Journal), 2013, 1008. http://digitalcommons.unl.edu/libphilprac/1008/ (accessed on 10 January 2016).
10. Enang, U.U. The use of International Standard Bibliographic Description (ISBD) and the quality of books published in Nigeria. *Lib. Phil. and Pract.* (e-journal), 2008, 176.
11. Massey, O. & O'Brien, A. Auditing cataloguing quality by random sampling. Loughborough University. 2000. Master degree dissertation.
12. Chapman, A. & Massey, O. A catalog quality audit tool. *Lib. and Inf. Res. News.,* 2002, **23**(82), 26-37.
13. Hewitt, J.A. Sample audit of cards from a university library catalog. *Coll. & Res. Lib.*, 1972, **33**(1), 24-7.
14. Mansor, Y. Issues in developing a cooperative cataloging program in Malaysia: an analysis of MARC records in three university libraries' OPAC

databases. University of Pittsburgh, Pittsburgh. 1999. PhD Thesis. 173p.

Contributors

Mr Chandrappa is a research student at the Department of Studies in Library and Information Science, University of Mysore, Mysuru. Presently, he is working at Mysore University Library, Mysuru, Karnataka.

Dr N.S. Harinarayana is working as Associate Professor at the Department of Studies in Library and Information Science, University of Mysore, Mysuru. He teaches the courses library automation, information retrieval, and library classification and content organisation. He had experience of working at Indira Gandhi National Open University, New Delhi as Assistant Librarian for six years. He is a recognised supervisor for MPhil and PhD courses of University of Mysore.