

## Qualifying title statements of library catalogue of the University of Jaffna

Ketheeswaren, S<sup>a</sup>, Charles, E. Y<sup>b</sup> and Chandrasekar, K<sup>c</sup>

<sup>a,c</sup>*Library, University of Jaffna, Sri Lanka*

<sup>b</sup>*Department of Computer Science, Faculty of Science, University of Jaffna,  
Sri Lanka*

<sup>a</sup>vskethees@univ.jfn.ac.lk

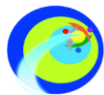
### Abstract

Metadata is the primary element for bibliographic communications. Hence, metadata should be qualified for a fruitful exchange of knowledge. The study examines the statements of title-proper and other-title-information of the monograph catalogue maintained by the University of Jaffna for quality status. The statements of title-proper and other-title-information have been taken from the same catalogue records for the study. Their quality status has been examined by comparing the respective statements maintained by the Library of Congress. The quality issues have been identified for the statements of the title-proper and the other-title-information, which have been contributed by leaving extra space, spelling mistake, typographical error, missing a part of the statement, missing a complete statement, the inclusion of extra part with the statement, and inclusion of complete extra statement. However, Typographical error and missing a complete statement are the most significant issues in qualifying title-proper and other-title-information, respectively. The catalogue records of the University of Jaffna do not have a substantial-quality issue for the bibliographic communication concerned through title-proper and other-title-information statements. However, the statements of title-proper and other-title-information of the University of Jaffna should be validated against qualifying the records to enhance bibliographic communication status.

*Keywords: cataloguing errors and catalogue quality*

### Introduction

The organisation of knowledge is a part of knowledge management. Information resources are the printed and digital media where knowledge is systematically codified. These information resources are organised by libraries using the techniques of cataloguing and classification (Igbini & Ikenwe, 2018). The catalogue for an information resource is prepared by transcribing metadata statements such as title, author, edition, publisher, copyright year, physical descriptions, etc. As a cataloguing management tool, Anglo-American Cataloguing Rules, Classified Catalogue Code are primarily used by the libraries. The rules stated in these tools explain how the punctuations and regulations should be used to make an understandable



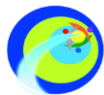
catalogue independent of the language of description and support the international exchange of bibliographic data. The catalogue of the Library of Congress is prepared strictly following the Anglo-American Cataloguing Rules (AACR) and regulations to avoid any kind of errors (Myall & Chambers, 2007; Shin, 2003). Hence, the LC catalogue could be used as a standard for evaluation studies.

The Library, University of Jaffna (UOJ) is also using the same rules (AACR) for transcribing the data elements of an information resource for preparing catalogues. Simultaneously, its library catalogue is originally prepared by the cataloguing librarians attached to its library. The originally prepared catalogues with human intervention are usually left with human-made errors. These errors and their distribution should be identified to correct them in existing catalogues and avoid them in future catalogues. The error-free catalogue will enhance the effectiveness of searching and retrieving bibliographic information from the Online Public Access Catalogue, and bibliographic communication of cooperative and union catalogues.

The comparison of UOJ records with the standard records would identify the issues on its quality. At this juncture, LC records can be selected in place of standard records. Further, examining the errors and their distributions in the title statement of information resources becomes foremost significant since the title statement is considered a mandatory element in the library catalogues. Hence, this study proposed identifying the categories of errors and their distribution appearing in the title statement of the library catalogue of the University of Jaffna and reasons for their existence to propose the methods of eliminating the same.

### **Literature Review**

Several research studies have looked into the problem of quality in cataloguing. The quality in cataloguing or metadata is defined in various research communications. Shin (2003) defines the quality in cataloguing as the assurance of extending and accuracy of the bibliography while it is defined as “dynamic and dependent on the values and needs of cataloguing users,” by Sarah (2005). Apart from defining the quality in cataloguing, authors also studied the different aspects of the quality in the catalogue: For example, Marc (2016) argued that the quality of the bibliographic details of the catalogue should assure the *completeness, accuracy, and consistency* in representing correct source datasets.



Some other studies were conducted on quality assurance to overcome inconsistent data storage, search, and retrieval issues. Moreover, the study carried out by David and Thomas (2015) correctly reflects on the scope of this present study, which discussed how the quality control procedures and workflow in checking the errors in the bibliographic records could be established.

*The importance of the metadata quality, evaluation of quality and methods for maintaining quality during cataloguing* were also discussed (Park, 2009; Park & Tosaka, 2010). *Cataloguing errors* can be usually categorised into two groups as *general* and *specific*, (Park, 2009). Further, the cataloguers' mistakes and approaches to maintain the standards in the catalogue metadata are also documented (Park, 2009). Lam (2007) conducted a study to identify the error rates in the bibliographic records of monographs prepared using copy cataloguing and locate where the *error rates* highly populated when conducting outsourcing program. In the same study, Lam (2007) compared two sets of samples, one from the OCLC cooperative cataloguing archive and another from the University of Saskatchewan's in-house cataloguing records. However, studies conducted by Bruce and Hillmann (2004); Ochoa and duval (2006, 2009) discussed the quality metrics as *Completeness, Accuracy, Confirmations to expectation, Logical Consistency and Coherence, Accessibility, Timeliness, and Provenance*.

Apart from studies on diverse aspects of *cataloguing quality* or *errors* as seen above, few studies focused on *specific errors*. Jeffrey and Karen (1995) particularly examined the *Typographical error* corrected during copy cataloguing. The study found that 35.8 percent of the errors had been corrected during copy cataloguing. Walker and Kulczak (2007) evaluated the original catalogue of the University of Arkansas Libraries for document retrieval issues. This study examined 298 titles, and results revealed that 99.66 % records correctly correspond to the respective books for physical verification, while 27.50 % of records do not correctly respond to retrieval of respective books. Besides, 38.59 % of records required modifications in the titles.

Concerning quality assurance of catalogue, the literature reviewed highlighted *types of errors, error distribution, quality level in the catalogues, quality improvement, and normalisation of error rates*.

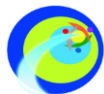


Table 1. Query for unmatched titles

---

```
SELECT 245a,245b,title  
FROM `UOJ-LC-Data-latest`  
LEFT JOIN `UOJ-Original-Data` on (isbn1=isbn)  
WHERE 245a != title
```

---

## Methodology

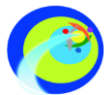
125,482 monograph catalogue records have been identified in the University of Jaffna (UOJ) Library system of 9<sup>th</sup> May 2020. The ISBNs of the records were selected as inputs to capture the corresponding records from the Library of Congress. Hence, the records with ISBN (n=52,705) have been selected for the study in the collection of records of the UOJ. Marccedit software was selected as a tool for harvesting records from the LC for the bulk ISBN inputs.

9,265 bibliographic records have been harvested from the Library of Congress (LC) for the inputs of 52705 ISBNs. However, multiple records for an ISBN input have been identified in the copied records (n=9,265) from the LC. After eliminating the ISBNs for which multiple records were retrieved, 7,235 numbers of ISBNs have been identified as appropriate inputs in retrieving a single record for a single ISBN from the LC. Out of 7,235 ISBN inputs, 5049 records have been retrieved from LC, while 2186 numbers of records have been reported lost due to random connection errors. Therefore, 5049 number of UOJ records would have corresponding records copied from LC. Hence, the total population of the study is 5049 in the number of records.

Subsequently, the records copied from the LC (n=5049) and records of ISBN inputs from UOJ (n=9,265) were managed in a MySQL database. Then, the query shown in Table 1 for collecting unmatched records from both collections was run to select the columns named 245a (titles) and 245b (other title information) from the table: UOJ-LC-Data-latest (records of LC), and “title” (title) from the table: UOJ-Original-Data (the records of UOJ). The selected column values were exported to an excel sheet. After that, investigations were made to identify the possible reasons for the unmatched conditions of the titles. The functions available in MS excel have been used to analyse the records.

## Results and Discussions

From the total population (n=5049), 3966 numbers of title statements have been identified as unmatched. In the 3966 numbers of title statements, title-proper has been found in each UOJ and LC record. However, in 1934, 1712



numbers of records have been found with other title information (OTI) in LC and UOJ, respectively. The records have been copied from two different library systems of LC and UOJ. Consequently, dissimilarity in copied records from both systems was observed. The dissimilarities among two sets of records should be observed in a unified condition. However, the punctuations have been exported within the records of LC while not exported within UOJ. Hence, the punctuations have been removed from the LC records to compare with the UOJ records.

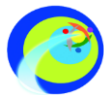
After removing the punctuations from the LC records, 1018 numbers of title-proper and 1929 numbers of other title information (OTI) have been left unmatched. This reveals 20.2% of title-proper and 38.2% of subtitles left with quality issues. Possible corrections have been identified for qualifying the title statements of the UOJ as detailed in table 2.

Table 2. Percentage of corrections against the total population

Field	Missing a part of statement	Typographical error	Inclusion of extra part with the statement	extra space	Spelling mistake	Missing a complete statement	Inclusion of complete
Title-proper	8.14%	4.44%	4.2 %	2%	0.7%	0%	0%
OTI	10.53 %	1.76%	0.71 %	17 %	6 %	7 %	2.5 %

The most profound issue in qualifying the title-proper is a portion of the statement is missing in the field (8.14%). The reason for this issue is that the UOJ library system has left the last parts of the lengthy titles during export of the records. Therefore, *missing a part of the data field's statement* cannot be taken as an essential issue in qualifying the records. In addition to this issue, leaving *extra space* have no impact on bibliographic communication since many information systems neglect the spaces during the indexing of data. Hence, the summation of percentage contributed by the other errors (9.34 %) such as *Typographical error*, *Inclusion of extra part with the statement*, *spelling mistake*, *Missing a complete statement* and *Inclusion of complete extra statement* would reflect on qualifying the statements of title-proper of the UOJ records as the overall issues.

Similarly, these errors contributed 17.43% in the statements of OTI. The quality issues identified for title-proper (9.34 %) and OTI (17.43%) would have considerably affected the effective bibliographic communication. The

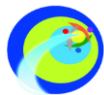


measures should be taken to enhance the UOJ records' quality by using systematic protocols, as discussed below.

An issue in qualifying the title-proper namely, *Typographical error*, took the value of 4.44%. The issue of typographical error can be rectified only by incorporating typing techniques during the data entry. 4.2% of corrections have been identified based on adding extra portions with the title-proper. The reason for the phenomenon is OTI has been added with title-proper treating OTI as complained of the title-proper. The meaningful inference of title-proper and OTI from the source can only rectify this issue. Either complete missing of a field or new extra field was not identified (0%) as an issue in the title-proper since the title's statement is treated as a mandatory field in the UOJ library system. In OTI, *extra space* is a profound issue (17 %) about the quality of the catalogue. The phenomena infer that OTIs have more word divisions than titles or cataloguers may be more inclined to create *extra space* in the latter part of a typing passion. In 10.53 % of records, portions of OTI are missing. The same are twofold, improper or alternative inference of metadata from the source or missing of lengthy portions during the export of records from the UOJ library system. Concentration should be drawn on inferring of OTI from the source to rectify this particular issue. Spelling corrections are identified as a quality issue in OTI records, though it enrolls lesser value for title-proper (0.7%). According to the UOJ library, application of spelling proof is not covered in the cataloguing guidelines, and this is the reason for the existing *spelling mistake* in the records. Further, 2.5% of issues on qualifying the OTI have arisen from the inclusion of unnecessary data in the field of OTI. The meaningful selection of OTI from the source will reduce this issue.

### Conclusions and Recommendations

The profound issue in qualifying the title-proper is *missing a part of entry* of data in the field; it is only due to the library system's error in exporting the record. Though leaving *extra space* has been taken as an issue in qualifying metadata's statement, it would not impact bibliographic communication since most indexing systems neglect additional spaces. However, the other quality issues appeared on title-proper and OTI, such as *Typographical error*, *Inclusion of extra part with the statement*, *spelling mistake*, *Missing a complete statement* and *Inclusion of a complete extra statement*, make considerable impacts on bibliographic communication. *Typographical error* is the most significant issue in qualifying title-proper, which has been identified in 4.44 % of records. At the same time, *missing a complete statement* is the major issue in OTI, which has been identified in 7 % of

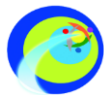


records. The *spelling mistake* has also been identified as a quality issue in OTI statements (6%), which should be avoided while feeding data into the library system. A spelling proof tool can be employed for the same by adding an extension for spelling proof into the internet browser. “Grammarly” is an example that can be added as an extension to the internet browser as a spelling proof tool. Techniques of typing and rules applied on the inference of title-proper and OTI should be systematically incorporated during the data entering to overcome the quality issues arisen from *Typographical error*, and the inclusion of *extra space* and incorrect inference of meaningful bibliographic data from the source.

The total contribution of errors on title-proper and OTI is 9.34 % and 17.43%, respectively. The readers usually execute searching by using a part of the title. Hence, these total contribution of errors does not impact on retrieving catalogued books in the identified percentage. Hence, the UOJ records do not have a serious quality issue supporting the search and retrieval of catalogued books using keywords that focus on Title-proper and OTI. However, the identified quality issues should be rectified by conducting validation to enhance bibliographic display and communication status, which would also be useful for supplying qualified records for cooperative and union catalogues. In addition to this, those issues should be avoided in future catalogues by revising the guidelines based on this study's findings. The records managed by the UOJ library system database are not populated with required punctuations in the database and the punctuations are automatically added by the library system only when displaying catalogue in the ISBD view. Hence, the punctuations were not included with the catalogue records exported from the UOJ library system. Hence, the applications of punctuations on the title-statements used by the UOJ library have not been possibly verified compared to the LC records could be considered a limitation of the study.

## References

- Bruce, T., & Hillmann, D. (2004). *The Continuum of Metadata Quality: Defining, Expressing, Exploiting* Vol. 1. *eCommons Metadata in Practice* (pp. 17). doi:[https://ecommons.cornell.edu/bitstream/handle/1813/7895/Bruce\\_Hillmann\\_corr\\_final.doc?sequence=1&isAllowed=y](https://ecommons.cornell.edu/bitstream/handle/1813/7895/Bruce_Hillmann_corr_final.doc?sequence=1&isAllowed=y)
- David, R. H., & Thomas, D. (2015). Assessing Metadata and Controlling Quality in Scholarly Ebooks. *Cataloging & Classification Quarterly*, 53(7), 801-824. doi:10.1080/01639374.2015.1018397



- Igbinovia, M., & Ikenwe, I. (2018). Knowledge management: processes and systems. *Information Impact: Journal of Information and Knowledge Management*, 8, 26. doi:10.4314/ijikm.v8i3.3
- Jeffrey, B., & Karen, K. (1995). The Effectiveness of Copy Cataloging at Eliminating Typographical errors in Shared Bibliographic Records. *LRTS*, 48(2), 92-100.
- Lam, V.T. (2007). Error Rates in Monograph Copy Cataloging Bibliographic Records Before and After Outsourcing at the University of Saskatchewan Library. *Cataloging & Classification Quarterly*, 44(3-4), 213-220. doi:10.1300/J104v44n03\_04
- Marc, D. T. (2016). *Assessing Metadata Quality and Terminology Coverage of a Federally Sponsored Health Data Repository*. (The Degree of Doctor Oo Philosophy), University of Minnesota, Minnesota.
- Myall, C., & Chambers, S. (2007). Copy Cataloging for Print and Video Monographs in Two Academic Libraries: A Case Study of Editing Required for Accuracy and Completeness. *Cataloging & Classification Quarterly*, 44(3-4), 233-257. doi:10.1300/J104v44n03\_06
- Ochoa, X., & duval, e. (2006). Towards Automatic Evaluation of Learning Object Metadata Quality. *Advances in Conceptual Modeling - Theory and Practice*, 4231(1), 372-381. doi:https://doi.org/10.1007/11908883\_4
- Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International Journal on Digital Libraries*, 10(2), 67-91. doi:10.1007/s00799-009-0054-4
- Park, J.-R. (2009). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly*, 47(3-4), 213-228. doi:10.1080/01639370902737240
- Park, J.-R., & Tosaka, Y. (2010). Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms. *Cataloging & Classification Quarterly*, 48(8), 696-715. doi:10.1080/01639374.2010.508711
- Sarah E, T. (2005). Quality in Bibliographic Control. *Library Trends*, 44(1996), 491-505.
- Shin, H.-S. (2003). Quality of Korean Cataloging Records in Shared Databases. *Cataloging & Classification Quarterly*, 36(1), 55-90. doi:10.1300/J104v36n01\_05
- Walker, M., & Kulczak, D. (2007). Shelf-ready books using PromptCat and YBP: Issues to consider (An analysis of errors at the University of Arkansas). *Library Collections, Acquisitions, & Technical Services*, 31(2), 61-84. doi:10.1080/14649055.2007.10766150