International Journal of Research in Library Science (IJRLS)

ISSN: 2455-104X

DOI: 10.26761/IJRLS.11.1.2025.1832

Volume 11, Issue 1 (January-March) 2025, Page: 116-130, Paper ID: IJRLS-1832

Received: 28 Dec. 2024 ; Accepted: 3 Feb. 2025 ; Published: 8 Feb. 2025

Error Identification and Classification in Bibliographic Records of INI in India: A Study

Pradeep Kumara B.[1]; N. S. Harinarayana[2]

Research Fellow, Dept. of Studies in Library & Info. Science, University of Mysore[1]; Senior Professor, Dept. of Studies in Library & Info. Science, University of Mysore, Manasagangothri Campus, Mysuru, India[2]

pradeepkumarbraju@gmail.com, ns.harinarayana@gmail.com

ABSTRACT

This paper investigates the identification and classification of errors in bibliographic records in Institutions of National Importance (INI) in India. Accurate bibliographic records are crucial for information retrieval in libraries and academic institutions. This study aims to provide a comprehensive investigation into identifying and classifying errors in bibliographic records across six categories of Institutions of National Importance (INI): All India Institute of Medical Sciences (AIIMs), Rishikesh (Medical), IIM Ahmedabad, IIM Jammu, IIM Kolkata, IIM Kozhikode, IIM Udaipur (IIMs), IIT Gandhinagar, IIT Goa, IIT Kanpur (IITs), IISER Mohali, IISER Bhopal (IISERs), MNIT Jaipur, NIT Rourkela (NITs), and School of Planning and Architecture (SPA), Bhopal (Planning). A quantitative method was employed where the population of English books stacked on the OPACs was 6,92,166 as of February 2020. The population was too large to handle and a sample size of 1536 was arrived using Krejcie and Morgan formula. Errors are analyzed under three primary classifications—format errors, content errors, and edit & input errors. The findings showed that, for accessibility status of OPACs, out of 128 institutions, more than half 70(54.69%) INI had non-availability of OPACs, 38 (29.69%) had inaccessible OPACs, 6(4.69%) OPACs without MARC data export, and 14 (10.94%) had usable OPACs. Similarly, INI indicated different distributions of types of errors in record categories. That is, in Medical Record, 300 fields had 152 (20.16%) followed by 245 tags with 144 (19.10%). In IIMs, the 300 field had 2274 (20.68%) and the 082 field had 2111 (19.20%) errors. For IISERs, the 082 fields had 157 (20.47%), and 300 fields had 145 (18.9%) errors. The results emphasize the need for targeted interventions, including AI-based tools and standardized cataloguing practices, to enhance bibliographic quality.

.

KEYWORDS: Bibliographic Record, Format Errors, Content Errors, Input Errors, Catalogue Quality.

INTRODUCTION

The widely circulated propositions attributed to Manfred Kochen (1976, p. 150) cited in Daniel, (1993, p. 664) that, “an information system may be used, but not be useful; it may also be useful, but not used. It may even be neither useful nor used. It is ideal if it is both used and useful” shed more light on why those systems have to make

themselves both used and useful to their end users. Perhaps this is one of the reasons why technical service departments in libraries aid in acting as an engine for ensuring the quality of services such as bibliographic records and their maintenance for easy access by users. This department does so by engaging in various processes that deal with cataloguing, classification, and indexing critical for organizing human knowledge. Catalogues are information retrieval tools that transform print to digital in the form of OPAC, then web-based OPAC, web OPAC, to online union catalogue whose goal is to warrant finding information from any institution at any location. From another perspective, different types of errors are prevalent in databases, which include but are not limited to "duplicate records, typographical errors, tagging errors, inconsistent headings of entries, and missing data" (Chandrappa & Harinarayana, (2018) cited in Zakaria, 2023, p. 849). This implies that, the supposedly universal coverage can easily be trumped by many problems including errors among which is a typographical error emanating from failure to critically adhere to cataloguing rules, inconsistent formatting, avoidable typos, to mention but a few. These errors prevent the retrieval of bibliographic records, which usually represented in an inverse relationship that, the higher the catalogue quality, the lower the occurrence of errors in records (Zakaria, 2023). That is why many researchers investigated the phenomenon of typographical errors for its relevance in the field of LIS profession. For instance, Zakaria, (2023) pointed out many researchers who have shown interest in this direction among whom is Ballard. Ballard was able to detect errors in OPACs thereby developing a systematic way of eliminating typographical errors from the 'Dirty Database Test'. From the list, errors can emerge from personal name, corporate name, conference name, title, and LC subject heading, among others.

From the above paragraph, it is evident that, the term quality is ambiguous as it bewilders researchers since it means different things to different researchers, context, content, time, and space. The problem of quality in terms of inaccuracy, obsolescence, duplication, triviality, etc. is not a new phenomenon in the LIS profession as Daniel (1993) cited William, (1993, p. 644) who categorically appreciated the growth of database records from 52 million in 1975 to 5 billion in 1989 but with consequences of circulating available databases that are problematic. She was able to demonstrate the "80/20 rule" or "vital few and trivial many" to justify how a user might be confronted with only a tiny fraction of useful information resources in the pool of unimportant or less important ones due to the problem of errors. That is why she calls for quality improvement in databases (Daniel, 1993). This means that quality is all-encompassing as it has a wider spectrum in terms of many of its attributes. This is true, as Chandrappa and Harinarayana, (2018) have pointed out many arguments as regards what constitutes quality and thus difficulty in discerning what could be catalogue quality. It is clear that catalogue quality, too has to have such inherent confusion as some scholars referred to it as abiding by cataloguing rules that ensure maintaining a clean database that demonstrates the integrity of the institution (Penkiunas, 1995 Cited in Chandrappa & Harinarayan, 2018). and others narrated other things differently.

Even though the quality of records in library databases is not as costly as financial documents; quality is critical for economic, social, environmental, medical, technological, etc. perspectives. For instance, Randall, (1999, p. 161) referenced Frost and Goldner (1999, p. 161) who captured a scenario where three zeroes eliminated from financial figures resulted in the loss of a significant amount of money in millions of dollars.

In the analysis of the study, Randall, (1999, p. 165) noted that “unique errors are not all equal”. For example, errors in a title or subject field are more serious than errors in a note” since users rely heavily on keywords and subjects to search for information. To concur with this fact, Randall cited Ballard and Lifshi, (1992) whose study indicated that errors were more prevalent in “title fields (63%), note fields (21%), author errors (9%), and series errors (7%)” (Randall, 1999, p. 165). Fortunately, her study found errors in five fields “author fields (including main author, alternate author, and publisher), title fields (including main and alternate titles), subject fields, note fields, and series fields”. That is why, part of her concluding remarks, she reiterated that, even though perfection is worthwhile in humans and databases, it is also impossible (Randall, 1999).

A critical review of the literature reveals the vital role of quality control in cataloguing and its broader influence on library services. Over the years, the quality of cataloguing has consistently been a focus of research and discussion. Avdoyan (1995) defined cataloguing quality as the consistent production of detailed bibliographic records that meet high research standards while remaining accessible to general users. This dual-purpose objective highlights the need for catalogue records to balance precision with ease of use. Graham (1990) differentiated between two essential aspects of accuracy in cataloguing. First, "mechanical accuracy" relates to typographical precision and proper transcription. Second, "intellectual accuracy" concerns how well bibliographic descriptions correspond to the actual items being catalogued. These elements together form the basis for effective and reliable information retrieval. From another angle, Taylor (1992) noted significant challenges in cataloguing personal name access points, pointing out how inconsistencies can hinder both search recall and precision. In his study, discrepancies in name headings were identified in 17.7% to 24.1% of records. Similarly, Mansor (2003) analyzed 410 MARC records and found that only 20.2% adhered completely to established cataloguing standards.

Importantly, Zeng (1993, 1994) made significant contributions to understanding cataloguing errors by categorizing them into three primary types: format errors, content errors, and editing/input errors. Format errors—such as incorrect field tags or missing subfield codes—create major issues for automated systems. Content errors, which involve missing or inaccurate bibliographic details, are particularly harmful to retrieval accuracy. Editing and input errors, though less severe, undermine the professional quality of cataloguing records. That is why Thomas (1996) emphasized that maintaining high cataloguing standards is crucial, even if users often accept partial search success. His research highlighted that even minor errors can accumulate over time, eroding user trust in catalogues and diminishing overall satisfaction with library services. Interestingly, the development of automated tools, such as validation systems, has opened up possibilities for improving cataloguing accuracy. Paiste (2003) argued that quality control in cataloguing should go beyond error elimination to focus on achieving consistency, depth, and timeliness in bibliographic records.

Studies on international databases like OCLC and RLIN have also provided benchmarks for cataloguing standards. Intner (1989) found that inconsistencies in cataloguing practices across these databases often result in discrepancies that complicate cross-library resource sharing. Zeng (1993) specifically studied Chinese-language records in OCLC’s database, identifying widespread errors in both system-generated and user-submitted records. These findings underscore the importance of standardization in cataloguing. To be precise, the literature makes a strong case for ongoing improvements in cataloguing practices. While technological advancements hold promise for

reducing errors, the human expertise required to uphold cataloguing standards and ensure intellectual accuracy remains irreplaceable. Building on these insights, this study examines cataloguing quality across INI libraries, offering recommendations to enhance standards and practices. Bibliographic records serve as the backbone of modern library systems, enabling efficient information organization and retrieval. However, inaccuracies in these records can significantly hinder user access to resources. This study explores errors in bibliographic records across six major institutional categories of INI: Medical, IIMs, IITs, IISERs, NITs, and Planning. By classifying errors into three categories—Format, Content, and Edit & Input Errors—this research identifies trends and inconsistencies, providing a comparative analysis and actionable insights for improved cataloguing practices. The purpose of the study is to identify and classify errors in bibliographic records of chosen libraries.

METHODOLOGY

This study systematically identifies and classifies errors in catalogues in libraries of Institutions of National Importance (INI) in India. It involves the collection of bibliographic data from publicly accessible OPACs, analysis of metadata adherence to standards like AACR2, RDA, and MARC21, and identification of cataloguing errors categorized into format, content, and edit & input errors. The population of the institutions is 128 and research collection from 14 selected institutions as of February 2020 was 6,92,166. The sample size is 1536 arrived at using Krejecie and Morgan Formula. The institutes were categorized into 6 categories: IIMs, IITs, Medical, IISERs, NITs, Planning. Data collection is a critical component of this study, designed to ensure accuracy, consistency, and representativeness of the data. The process involved retrieving cataloguing records in MARC21 format from the selected 14 Institutions of National Importance (INIs). By leveraging Online Public Access Catalogues (OPACs) and institutional systems, the study ensured the data was comprehensive and suitable for evaluating cataloguing practices. Procedure for data analysis involves

- Institutions were contacted to obtain permissions for accessing OPAC data where necessary.
- Assistance was sought from library staff for exporting records in MARC21 format.
- Web OPACs were used to directly download cataloguing records.
- Records were exported in **ISO 2709 format**, ensuring compatibility with cataloguing analysis tools.
- MARC Edit was utilized to convert data from the ISO 2709 format into Microsoft Excel.
- Subsequent processing of the data was primarily carried out using Microsoft Excel.
- Only records in the **English language** were included in the sample to maintain consistency.
- Duplicates, incomplete entries, and irrelevant records were removed during the data cleaning process

FINDINGS

Table 1 Error type codes

Sl. No	Type of Error	Code	Description
1.	FORMAT ERRORS	F1	Incorrect field tag
		F2	Incorrect Indicator
		F3	Omitted subfield code; Incorrect subfield code; Incorrect sequence of subfields.
		F4	Other format differences

Sl. No	Type of Error	Code	Description
2.	CONTENT ERRORS	C1	Omitting a whole field; or containing an extra field
		C2	Omitting a part of the entry such as a subfield; or containing an extra part for the entry
		C3	Inconsistency between corresponding fields or value
		C4	Incorrect content - whole field
		C5	Incorrect content - subfield
		C6	Other content errors
3.	EDITING & INPUTTING ERRORS	E1	Spacing
		E2	Misspelling
		E3	Capitalization
		E4	Comma
		E5	Colon
		E6	Period missing
		E7	Omitted slash
		E8	Semicolon

Table 1 shows error type codes. This format adopted from Chandrappa's (2019) thesis who cited Zeng, (1993) that, there are three classes of errors. The first is Format Error, which has four codes. Secondly, Content Error that has six codes, and thirdly, Edit and Input Error that has eight (codes). A field with three typographical errors was counted as three separate errors. However, these would not have significant impact on the quality of the record as a missing entry.

Table 2 Accessibility Status of OPACs in INIs in February 2020

Category	Number of Institutions	Description
Non-availability of OPACs	70 (54.69%)	Institutions without any publicly accessible OPACs.
Inaccessible OPACs	38 (29.69%)	Institutions with technical issues such as broken links or outdated platforms.
OPACs without MARC data export	6 (4.69%)	Institutions whose OPACs did not support MARC or ISO 2709 data export formats.
Usable OPACs	14 (10.94%)	Institutions with accessible OPACs and MARC21-compliant data for export.
Total	**128 (100%)**	

Table 2 shows the accessibility status of OPACs in INI as observed in February 2020. From the table, it can be seen that, institutions without any publically accessible OPACs accounted for 70 (54.69%) followed by inaccessible OPACs 38 (29.69%), usable OPACs 14 (10.94%) and OPACs without MARC data export 6 (4.69%). This implies

that, INIs have many institutions without OPACs, which can limit the accessibility of the contents by the end users. In addition, OPACs without MARC data export means that, those institutions have data that cannot be downloaded for retrieval purposes.

Table 3 Final Selection of Institutions

Category	Institutions
Medical	All India Institute of Medical Sciences (AIIMS), Rishikesh
IIMs	IIM Ahmedabad, IIM Jammu, IIM Kolkata, IIM Kozhikode, IIM Udaipur
IISERs	IISER Mohali, IISER Bhopal
IITs	IIT Gandhinagar, IIT Goa, IIT Kanpur
NITs	MNIT Jaipur, NIT Rourkela
Planning	School of Planning and Architecture (SPA), Bhopal

Table 3 illustrates the institutions chosen for their ability to provide MARC21-compliant data and its representation of a specific academic focus, ensuring that the study captures a comprehensive range of cataloguing practices. The rationale for choosing these institutions are many but can be due to their diverse academic disciplines, from management and engineering to medical and planning studies, adhering to MARC21 standards for ensuring uniformity in cataloguing data, have accessible and export-capable OPACs to facilitate seamless data collection and analysis.

Table 4 Proportional Distribution of Sample Records

Category	Total Population	English Book	Percentage (n=305375)	No. of Unique Records Considered for this Study	Sample Size
Medical	4641	4183	1.36979124	4	26 (1.69%)
IIMs	326282	138818	45.45820712	139	695 (45.25%)
IISERs	14226	11419	3.739336881	11	58 (3.78%)
IITs	228494	32665	10.6966844	33	165 (10.74%)
NITs	109888	109745	35.93778142	110	549 (35.74%)
Planning	8635	8545	2.798198936	9	43 (2.80%)
Total	692166	305375	100	305	1536 (100%)

Table 4 represents a proportional distribution of sample records drawn from six different categories of institutions or domains (e.g., Medical, IIMs, IISERs, IITs, NITs, and Planning) based on their total population. The idea is to allocate the sample size proportionally to the size of each category's total population. The sample size of English records against each category is derived by taking the round number of unique records. For instance, 26 is obtained by taking 1% of English book (based on the collections) then rounded up to the nearest whole number. In this direction, we took 50% of each of the record thereby approximating them to the nearest whole numbers as shown in the table.

Table 5 Error Distribution of Records—Categorized Based on Field

Fields	Errors in IIMs Records	Errors in IISERs Records	Errors in Medical Records	Errors in IITs Records	Errors In NITs Records	Errors in Planning Records	Total Errors	% Errors
020	566	47	17	139	401	68	1238	4.27
082	2111	157	104	617	2034	172	5195	17.93
100	1401	63	69	457	1369	150	3509	12.11
245	1247	132	144	911	1765	114	4313	14.89
250	183	19	46	96	137	12	493	1.70
260	1345	64	123	486	1201	83	3302	11.40
300	2274	145	152	422	2262	198	5453	18.82
440-490	117	5	0	118	141	17	398	1.37
500	105	20	26	94	121	11	377	1.30
504	506	32	20	126	317	37	1038	3.58
505	215	12	11	52	166	44	500	1.73
600	22	0	0	14	717	0	753	2.60
700	903	71	42	231	1101	52	2400	8.28
Total	**10995**	**767**	**754**	**3763**	**11732**	**958**	**28969**	100.00

Table 5 shows the distribution of errors categorized in field institution-wise. As it can be seen, the field 300 has the highest percentage of errors 18.82%. This implies that, most cataloguers in these institutions commit errors in Physical Description of information stocked in the websites. Furthermore, 082 (17.93%), which is Dewy Decimal Classification Number, has the second highest error encountered on the websites of catalogue records. In addition, Title Statement 245 (14.89%) followed by 100 (12.11%) Main Entry Personal Name, and 260 (11.40%) Publication Distribution Details occupied the next ranks. From another viewpoint, it can be observed that, those institutions with higher records have more errors. For instance, based on the analysis of this study, NITs have higher number of records and thus accounted for 11732 errors, which is not satisfactory result. This trend is observed in IIMs with 10995 errors in the institutions' records.

……..

Table 6 Distribution of Types of Errors in Medical Catalogue Records

Key: Numbers in brackets indicate percentages (%)

	Fields	020	082	100	245	250	260	300	440-490	500	504	505	700	Total (%)
FORMAT	F1	0	0	2	0	0	0	0	0	0	0	0	0	2 (0.27)
	F2	0	26	24	25	0	0	0	0	0	0	0	1	76 (10.8)
	F3	17	26	3	26	0	20	26	0	0	0	0	4	122 (16.18)
	F4	0	0	1	0	0	0	0	0	0	0	0	4	5 (0.66)
CONTENT ERRORS	C1	0	0	0	0	0	0	0	0	1	19	11	10	41 (5.43)
	C2	0	26	3	26	0	0	25	0	0	0	0	1	81 (10.74)
	C3	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	C4	0	0	4	0	26	0	0	0	0	1	0	4	35 (4.64)
	C5	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	C6	0	0	0	0	0	0	26	0	0	0	0	0	26 (3.44)
EDIT & INPUT ERRORS	E1	0	26	10	25	0	26	0	0	0	0	0	6	93 (12.33)
	E2	0	0	0	0	0	0	0	0	0	0	0	1	1 (0.13)
	E3	0	0	0	1	0	0	0	0	0	0	0	1	2 (0.26)
	E4	0	0	4	0	0	25	0	0	0	0	0	1	30 (3.97)
	E5	0	0	0	1	0	26	23	0	0	0	0	0	50 (6.63)
	E6	0	0	18	26	20	26	26	0	25	0	0	9	150 (19.89)
	E7	0	0	0	14	0	0	0	0	0	0	0	0	14 (1.85)
	E8	0	0	0	0	0	0	26	0	0	0	0	0	26 (3.44)
Total Errors %		17 (2.25)	104 (13.79)	69 (9.15)	144 (19.10)	46 (6.10)	123 (61.31)	152 (20.16)	0 (0)	26 (3.45)	20 (2.65)	11 (1.46)	42 (5.75)	754 (100)

Table 6 illustrates the distribution of errors across various MARC21 fields in records from the Medical. From the analysis, field 300 Physical Description has the highest number of errors with 152 (20.16%) followed by immediately by 245 Title Statement with errors of 144 (19.10%), followed by 260 Publication Distribution with 123 errors (16.31%). These findings suggest that key descriptive fields, such as titles and physical descriptions, are prone to frequent errors. Errors in these fields can significantly impact the discoverability and accurate representation of library resources. The field with the least number of errors is 440-490 Series Statements with no recorded errors (0%). This indicates either high accuracy in cataloguing this field or limited use of these fields across the records.

Similarly, field 440-490 has the minimal number of errors, accounting for only 0errors (3.45%), suggesting better consistency in input or fewer instances requiring this field.

Table 7 Distribution of Types of Errors in IIMs Catalogue Records

	Fields	020	082	100	245	250	260	300	440-490	500	504	505	600	700	Total (%)
FORMAT	F1	0	0	42	0	0	110	0	0	0	0	0	0	0	152 (1.38)
	F2	0	529	55	417	0	0	0	19	0	0	0	0	184	1204 (11)
	F3	565	445	104	63	0	34	456	0	0	0	0	0	1	1668 (15.17)
	F4	0	0	49	3	1	0	0	0	1	0	0	0	13	67 (0.61)
CONTENT ERRORS	C1	0	8	5	0	58	0	4	98	95	142	215	22	132	779 (7.09)
	C2	0	443	100	63	0	0	376	0	0	1	0	0	2	985 (8.96)
	C3	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	C4	0	0	1	1	4	0	5	0	0	352	0	0	2	365 (3.32)
	C5	0	0	1	0	3	0	8	0	0	0	0	0	0	12 (0.11)
	C6	0	0	117	3	0	0	20	0	2	9	0	0	74	225 (2.05)
EDIT & INPUT ERRORS	E1	1	686	248	87	0	675	358	0	0	0	0	0	190	2245 (20.42)
	E2	0	0	8	2	0	1	0	0	0	0	0	0	5	16 (0.15)
	E3	0	0	0	6	12	0	0	0	0	0	0	0	0	18 (0.16)
	E4	0	0	220	1	0	39	4	0	1	0	0	0	71	336 (3.06)
	E5	0	0	0	0	0	50	358	0	0	0	0	0	0	408 (3.71)
	E6	0	0	451	539	105	430	214	0	6	2	0	0	229	1976 (18.0)
	E7	0	0	0	62	0	0	1	0	0	0	0	0	0	63 (0.57)
	E8	0	0	0	0	0	6	470	0	0	0	0	0	0	476 (4.33)
Total Errors (%)		566 (5.15)	2111 (19.20)	1401 (12.74)	1247 (11.34)	183 (1.66)	1345 (12.23)	2274 (20.68)	117 (1.06)	105 (0.95)	506 (4.60)	215 (1.96)	22 (0.20)	903 (8.21)	**10995 (100)**

Key: Numbers in brackets indicate percentages (%)

Table 7 shows the distribution of errors in IIMs. From this table, field 300 Physical Description has the highest number of errors accounting for 2274(20.68%) followed by 082 DDC Number with number of errors 2111 (19.20%), followed by 100 Main Entry Personal Name accounting 1401 (12.74) and 260 Publication Distribution has 1345 (12.23%). This table shows the fields that received the maximum and minimum number of errors with 300 and 082 having the highest incidence of errors in the records. This means that, the physical description of information resources is not properly done, which can affect the retrievability and accreditation of those resources. The field that has the least number of errors is 600 Subject Added Entry Field with a number of error 22 (0.20%).

From this finding, it can be argued that, the least error encountered in the 600 Subject Access Fields is attributable to the fact that, most institutions do not incorporate such fields.

Table 8 Distribution of Types of Errors in IISERs Catalogue Records

	Fiel ds	020	082	100	245	250	260	300	440-490	500	504	505	700	Total (%)
FORMAT	F1	0	0	1	0	0	3	0	0	0	0	0	0	4 (0.52)
	F2	0	39	18	37	0	0	0	5	0	0	0	17	116 (15.12)
	F3	44	14	0	5	0	0	21	0	0	0	0	0	84 (11)
	F4	0	0	0	0	1	0	0	0	0	0	0	4	5 (0.65)
CONTENT ERRORS	C1	2	1	0	0	9	0	0	0	0	32	12	16	72 (9.39)
	C2	0	31	0	2	0	0	13	0	0	0	0	0	46 (6)
	C3	0	14	0	0	0	0	0	0	0	0	0	0	14 (1.83)
	C4	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	C5	0	0	0	1	0	0	0	0	0	0	0	8	9 (1.17)
	C6	0	0	14	2	4	0	2	0	0	0	0	7	29 (3.78)
EDIT & INPUT ERRORS	E1	1	58	19	39	0	54	56	0	1	0	0	13	241 (31.4)
	E2	0	0	0	1	0	0	0	0	0	0	0	0	1 (0.13)
	E3	0	0	0	5	0	0	0	0	0	0	0	0	5 (0.65)
	E4	0	0	2	2	0	1	0	0	0	0	0	2	7 (0.91)
	E5	0	0	0	2	1	4	21	0	0	0	0	0	28 (3.65)
	E6	0	0	9	19	4	1	9	0	19	0	0	4	65 (8.47)
	E7	0	0	0	17	0	1	1	0	0	0	0	0	19 (2.48)
	E8	0	0	0	0	0	0	22	0	0	0	0	0	22 (2.87)
Total Errors (%)		47 (6.13)	157 (20.47)	63 (8.21)	132 (17.21)	19 (2.47)	64 (8.34)	145 (18.9)	5 (0.65)	20 (2.61)	32 (4.17)	12 (1.56)	71 (9.26)	767 (100)

Key: Numbers in brackets indicate percentages (%)

Table 8illustrates the distribution of errors in IISERs. From this table, it is obvious that, the field 082 DDC Number has the highest number of errors accounting for 157(20.47%) followed by field 300 Physical Description with errors accounting for 145 (18.9%) followed by 245 Title Statement with number of errors 132 (17.21%) followed by 260 Publication, Distribution (Imprint) accounting for 64 (8.34%) and 100 Main Entry has 63 (8.21%). This table shows the fields that received maximum and minimum number of errors with 082 and 300 have the highest incidence of errors in the records. This means that, the DDC Number and physical description of information resources are not appropriately captured, which can affect the irretrievability and accreditation of those resources. The field that has the least number of errors is 440-490 Series Statement with number of error 5 (0.65%). From this finding, it can be argued that, the least error encountered in 440-490 Series Statement Field is attributable to the fact that, most institutions do not have enough journals captured on the institutions' fields.

Table 9 Distribution of Types of Errors in IITs Catalogue Records

	Fields	020	082	100	245	250	260	300	440-490	500	504	505	600	700	Total (%)
FORMAT	F1	0	0	14	0	0	0	0	0	0	0	0	0	0	14 (0.37)
	F2	0	151	134	148	0	1	0	8	0	0	0	2	51	495 (13.2)
	F3	137	165	46	100	25	51	45	42	22	1	0	2	26	662 (17.6
	F4	0	0	0	2	0	0	0	0	0	0	0	0	12	14 (0.37)
CONTENT ERRORS	C1	1	0	0	1	16	0	2	21	7	117	52	10	34	261 (6.94
	C2	0	136	59	83	14	0	30	31	25	0	0	0	25	403 (10.71)
	C3	0	0	1	4	0	0	0	0	0	0	0	0	0	5 (0.13
	C4	1	0	5	0	12	0	0	0	37	2	0	0	0	57 (1.51)
	C5	0	0	0	1	0	0	0	0	0	1	0	0	0	2 (0.05)
	C6	0	0	7	88	18	23	0	0	0	0	0	0	6	142 (3.77)
EDIT & INPUT ERRORS	E1	0	165	9	163	0	95	148	4	2	1	0	0	7	594 (15.8)
	E2	0	0	5	2	0	0	0	0	0	0	0	0	1	8 (0.21)
	E3	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	E4	0	0	95	15	0	102	65	0	0	0	0	0	30	307 (8.16)
	E5	0	0	0	7	0	100	29	0	0	0	0	0	0	136 (3.61
	E6	0	0	81	136	11	113	63	12	1	4	0	0	39	460 (12.22)
	E7	0	0	1	161	0	0	0	0	0	0	0	0	0	162 (4.31)
	E8	0	0	0	0	0	1	40	0	0	0	0	0	0	41 (1.09)
Total Errors (%)		139 (3.69)	617 (16.4)	457 (12.1)	911 (24.2)	96 (2.55)	486 (12.9)	422 (11.2)	118 (3.14)	94 (2.5)	126 (3.35)	52 (1.38)	14 (0.37)	231 (6.14)	3763 (100)

Key: Numbers in brackets indicate percentages (%)

Table 9 shows the distribution of errors in IITs. From this table, it is obvious that, the field 245 Title Statement has the highest number of errors 911 (24.2%) followed by 082 DDC Number with 617 (16.4%) followed by 260 Publication, Distribution, (Imprint) 486 (12.9%) followed by 100 Main Entry 457 (12.1%), 300 Physical Description with errors accounting for 422 (11.2%). This table shows the fields that received maximum and minimum number of errors with 245 and 082 have the highest incidence of errors in the records. This means that, the title, which users who mostly search information resources using title and DDC Number physical description of information resources are not appropriately captured, which can affect the retrievability and accreditation of those resources. The field that has the least number of errors is 600 Subject Added Entry with number of error 14 (0.37%). From this finding, it can be argued that, the least error encountered in 600 Subject Added Entry Field is attributable to the fact that, most institutions do not include subject added entries on the institutions' fields.

Table 10 Distribution of Types of Errors in NITs Catalogue Records

	Fields	020	082	100	245	250	260	300	440-490	500	504	505	600	700	Total (%)
FORMAT	F1	0	0	43	0	0	163	0	0	0	0	0	0	1	207 (1.76)
	F2	0	397	284	338	0	51	1	97	0	0	8	178	324	1678 (14.3)
	F3	365	543	70	246	0	1	427	1	0	0	0	178	23	1854 (15.8)
	F4	0	0	9	0	0	0	0	0	0	0	0	0	2	11 (0.1)
CONTENT ERRORS	C1	31	4	11	0	43	8	31	34	116	315	149	4	114	860 (7.3)
	C2	0	545	92	243	0	1	428	2	1	0	2	178	32	1524 (12.99)
	C3	1	0	0	0	0	0	0	0	0	0	0	0	0	1 (0)
	C4	0	0	42	0	1	0	0	0	0	0	0	0	2	45 (0.4)
	C5	0	0	163	1	0	0	0	0	0	0	0	1	143	308 (2.6)
	C6	3	0	2	17	27	6	3	0	0	0	2	0	2	62 (0.5)
EDIT & INPUT ERRORS	E1	1	545	94	166	0	500	428	0	3	0	0	0	50	1787 (15.23)
	E2	0	0	6	22	15	23	0	0	0	0	0	0	11	77 (0.7)
	E3	0	0	42	0	0	20	0	7	0	0	0	0	30	99 (0.8)
	E4	0	0	100	0	0	148	142	0	0	0	0	0	71	461 (3.9)
	E5	0	0	126	32	0	150	286	0	0	0	0	0	99	693 (5.9)
	E6	0	0	282	456	51	130	87	0	1	2	5	178	197	1389 (11.84)
	E7	0	0	3	244	0	0	0	0	0	0	0	0	0	247 (2.1)
	E8	0	0	0	0	0	0	429	0	0	0	0	0	0	429 (3.7)
Total Errors (%)		401 (3.42)	2034 (17.34)	1369 (11.67)	1765 (15.04)	137 (1.17)	1201 (10.24)	2262 (19.28)	141 (1.21)	121 1.03)	317 (2.7)	166 (1.41)	717 (6.11)	1101 (9.38)	11732 (100)

Note. Numbers in parentheses represent percentages.

Table 10 shows the distribution of errors in NITs. In this table, it is crystal clearthat, the field 300 Physical Description has the highest number of errors accounting for 2262 (19.2%) followed by 082 DDC Number 2034 (17.34%) followed by 245 Title Statement 1765 (15.04%), 100 Main Entry 1369 (11.67%), 260 Publication, Distribution, etc. (Imprint) 1201 (10.24%). The field that has the least number of errors is 500 General Note with number of error 121 (1.03%). From this finding, it can be argued that, the least error encountered in 500 General Note Field is attributable to the fact that, most institutions do not include general note on the institutions' fields.

Table 11 Distribution of Types of Errors in Planning Catalogue Records

	Fields	020	082	100	245	250	250	300	440-4490	500	504	505	600	700	Total (%)
FORMAT	F1	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	F2	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	F3	28	0	0	0	0	0	0	0	0	0	0	0	0	28 (16)

	F4	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
CONTENT ERRORS	C1	1	8	1	0	0	0	0	0	3	0	24	0	2	39 (20)
	C2	0	0	0	1	0	0	1	0	0	0	0	0	0	2 (1.2)
	C3	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	C4	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	C5	0	0	0	0	0	0	0	0	0	0	0	0	1	1 (0.6)
	C6	0	0	0	0	0	0	6	0	0	0	0	0	0	6 (3.5)
EDIT & INPUT ERRORS	E1	5	35	28	1	0	9	1	0	0	0	0	0	15	94 (55)
	E2	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	E3	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	E4	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	E5	0	0	0	0	0	0	1	0	0	0	0	0	0	1 (0.6)
	E6	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
	E7	0	0	0	1	0	0	0	0	0	0	0	0	0	1 (0.6)
	E8	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)
Total Errors (%)		34 (20)	43 (25)	29 (17)	3 (1.7)	0 (0)	9 (5.2)	9 (5.2)	0 (0)	3 (1.7)	0 (0)	24 (14)	0 (0)	18 (10)	172 (100)

Note. Numbers in parentheses represent percentages.

Table 11 shows the distribution of errors in Planning Institutions. In this table, it is crystal clear that, the field 082 DDC Number has the highest errors in the record accounting for 43(25%) followed by 020 International Standard Book Number 34(20%), 100 Main Entry 29(17%), 505 Formatted Content Note 24(14%), 700 Added Entry-Personal Name 18(10%). The records with least errors were 250 Edition Statement 0(0%), 440-490 Series Statement 0(0%), 504 Bibliography Note 0(0%), 600 Subject Added Entry 0(0%).

Comparison of Errors in Six INIs Catalogue Records

Table 12 Comparison of Distribution of Errors in the Catalogue Records of Six Category Libraries

Error Type	**Medical**	**IIMs**	**IISERs**	**IITs**	**NITs**	**Planning**	**Total**	**(%)**
Format Error	205	3091	209	1185	3750	28	8468	30.05
Content Error	183	2366	170	870	2800	48	6437	22.84
Edit & Input Error	366	5538	388	1708	5182	96	13278	47.11
Total	754	10995	767	3763	11732	172	28183	100

Table 12 shows that, the Edit & Input Errors account for the highest number of errors across all library categories, totalling 13,278 (47.11%). This is followed by Format Errors, with 8,468 errors (30.05%), and Content Errors, with 6,437 errors (22.84%). The data suggests that manual data entry and editing processes contribute significantly to cataloguing inaccuracies, highlighting the need for improved training and automation in these areas. Among the six categories of libraries, NITs (National Institutes of Technology) recorded the highest number of total errors, with 11,732 (41.62%), followed by IIMs (Indian Institutes of Management), with 10,995 errors (39.02%). These two

categories dominate the overall error count, indicating systemic issues in their cataloguing practices. Conversely, Planning Libraries reported the lowest total errors, with just 172 (0.61%), reflecting better cataloguing practices or fewer cataloguing activities. Similarly, Medical Libraries reported a comparatively low error count of 754 (2.68%), suggesting relatively better accuracy in cataloguing.

DISCUSSION

From the findings of this study, it is obvious that, errors occurring in records are not similar across Institutions of National Importance for varying reasons. This suggests that, despite efforts made to remove errors in those records, they are still rampant in catalogues. The presence of these errors in these institutions raises concerns as to how they limit the accessibility and discoverability of information resources by users. From Table 1, it is obvious that, when it comes to describing content in information resources, cataloguers commit a myriad of errors. Fortunately, previous and contemporary experts in the field such as Zeng, (1992, 1993, 1994) have identified types of errors in OPACs of institutions. Fundamentally, there are three (3) types of errors namely format errors, content errors, and edit and input errors with codes F1-F4, C1-C6, and E1-E8 respectively. Based on these errors, many tables were constructed. For each institution, library website was consulted and the English books downloaded were categorized into these types of errors in MS Excel, checked and compared with similar LoC catalogue records. From this comparison, accessibility of six category institutions' records was compared where some institutions have no OPACs 70(54.69%), inaccessible records 38(29.69%), OPAC without data export 6(4.69%), and usable OPACs 14(10.94%). These usable OPACs offered a promising foundation to conduct this research as contained in Table 2. These institutions are captured in Table 3 serving as the final selection of institutions, which are Medical, IIMs, IISERs, IITs, NITs, and Planning. Table 4 provides proportional distribution of sample records and Table 5 indicates the distribution of records catalogued based on fields. This implies the need to improve training of cataloguers on the frequency of errors attributable to incoherent alignment of practice and theory. This suggests taking steps further to augment manual cataloguing processes with automation in these areas especially through the use of emerging technologies such as AI to reduce presence of these errors in those records. This speeds up retrieval of information with accuracy and precision.

CONCLUSION

Presence of errors in INIs catalogue records is a serious problem for the management, users, and information professionals. One of the fundamental observations made in this study is that, the smaller the sample size, the larger the number of errors. This suggests that, all hands should be on deck to checkmate the entries made by cataloguers and ensure the use of technology appropriately to enhance the cleaning process of this practice.

REFERENCES

[1] Avdoyan, L. (1995). Cataloging quality: A Library of Congress symposium. In *Quality cataloguing*, *6*. Library of Congress.Chandrappa. (2019). *An assessment of cataloguing quality in University libraries in Karnataka* [University of Mysore]. Unpublished PhD thesis submitted to the University of Mysore in partial fulfillment for the award of PhD in Library and Information Science.
https://shodhganga.inflibnet.ac.in:8443/jspui/handle/10603/323259

[2] Chandrappa, & Harinarayan, N. S. (2018). Quality of cataloguing records of university libraries in Karnataka: A study. *Gyankosh: The Journal of Library and Information Management*, *9*(2), Article 2. https://doi.org/10.5958/2249-3182.2018.00007.2

[3] Chandrappa, & Harinarayana, N. S. (2018). What Constitutes Cataloguing Quality? A Review of Issues and Perceptions. *ILIS Journal of Librarianship and Informatics*, *1*(1), 83–93.

[4] Daniel, E. (1993). *Quality Control of Documents*. *41*(4), 644–664.

[5] Graham, P. S. (1990). Quality in cataloging: Making distinctions. *Journal of Academic Librarianship*, *16*(4), Article 4.

[6] Intner, S. (1989). Quality in bibliographic databases: An analysis of member-contributed cataloging in OCLC and RLIN. *Advances in Library Administration and Organization*, *8*, 1–24.

[7] Mansor, Y. (2003). Bibliographic exchange in Malaysia: Variations in name headings. *Library Review*, *52*(1), 38–42.

[8] Penkiunas, R. M. (1995). Cataloging quality: A Library of Congress symposium. In *Quality cataloguing* (Vol. 6, p. 31). Library of Congress.

[9] Randall, B. N. (1999). Spelling Errors in the Database: Shadow or Substance? *Library Resources & Technical Services*, *43*(3), Article 3. https://doi.org/10.5860/lrts.43n3.161

[10] Starr Paiste, M. (2003). Defining and achieving quality in cataloging in academic libraries: A literature review. *Library Collections, Acquisitions, & Technical Services*, *27*(3), 327–338. https://doi.org/10.1080/14649055.2003.10765935

[11] Taylor, A. G. (1992). Variations in personal name access points in OCLC bibliographic records. *Variations in Personal Name Access Points in OCLC Bibliographic Records*, *36*(2), 224–241.

[12] Thomas, S. E. (1996). Quality in Bibliographic Control. *Library Trends*, *44*(3), 491–505.

[13] Zakaria, M. S. (2023). Measuring Typographical Errors in Online Catalogs of Academic Libraries Using Ballard's List: A Case Study from Egypt. *Cataloging & Classification Quarterly*, *61*(7–8), 848–870. https://doi.org/10.1080/01639374.2023.2282579

[14] Zeng, L. (1992). *An evaluation of the quality of Chinese-language records in the OCLC OLUC database and a study of a rule-based data validation system for online Chinese cataloging*. https://catalog.hathitrust.org/Record/002706828

[15] Zeng, L. (1993). 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. *Cataloging & Classification Quarterly*, *16*(4), Article 4. https://doi.org/10.1300/J104v16n04_03

[16] Zeng, L. (1994). 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. *Cataloging & Classification Quarterly*, *18*(1), 3–26. https://doi.org/10.1300/J104v18n01_02