

Cultural Archives in the National Digital Library of India: An Analytical Study

A.Muthuraj *Nalanda Knowledge Resource Centre, Ganpat University Goa, South Goa, Goa, India.*muthurajmphilbdu@yahoo.com

RESEARCH ARTICLE

Open Access

ARTICLE INFORMATION

Received: March 20, 2026

Accepted: August 11, 2026

Published Online: September 01, 2026

Keywords:

Digital Archiving, Cultural Archiving, National Digital Library of India, Metadata

ABSTRACT

Background: This paper analyses the cultural archive resources available in the National Digital Library of India (NDLI). Digital archiving involves the preservation, organization, and long-term accessibility of digital resources, while cultural archiving focuses on preserving historical records, manuscripts, artworks, photographs, audiovisual materials, ritual practices, and other forms of cultural heritage. The National Digital Library of India (NDLI), developed by the Indian Institute of Technology Kharagpur and sponsored by the Ministry of Education, Government of India, serves as a national digital repository providing access to educational, research, cultural, and heritage resources through a single-window platform.

Purpose: The study analyses the distribution of cultural archive resources available in NDLI by examining content providers, archival content types, and language-wise collections. It further evaluates the concentration, variability, and inequality of digital cultural archive resources using quantitative analytical measures.

Methods: Cultural archive metadata were collected from the National Digital Library of India on 20 December 2025. The data were processed and analysed using Microsoft Excel 2021. Descriptive statistical techniques, including frequency, percentage, ranking, mean, median, standard deviation, and range, were employed. To provide a systematic analytical framework, the Coefficient of Variation (CV), Herfindahl–Hirschman Index (HHI), and Gini Coefficient were calculated to evaluate the variability, concentration, and inequality in the distribution of digital cultural archive resources.

Results: The study identified 78 content providers, 74 archival content types, 31 Indian languages, and 33 foreign languages represented in the NDLI cultural archive collections. The analytical results indicate that the repository exhibits high variability, moderate concentration, and substantial inequality in the distribution of digital resources. A small number of major international repositories contribute the majority of the collections, while Indian repositories play a significant role in preserving regional and indigenous cultural heritage. Books dominate the archival collections, whereas multimedia and emerging digital archival formats remain comparatively underrepresented. The language-wise analysis further reveals that a limited number of major Indian and foreign languages account for most of the archival resources, while several regional and minority languages have relatively limited representation.

Conclusions: The analysis highlights the content providers, types of resources, and language availability of cultural archives in the NDLI repository.

DOI: [10.15415/iie.2026.141002](https://doi.org/10.15415/iie.2026.141002)

1. Introduction

Archives are essential for preserving historical knowledge, supporting research, ensuring transparency, and maintaining institutional memory. Archive materials, such as manuscripts, photographs, maps, audio-visual items, and digital files, serve as evidence of the functions, decisions, and activities of individuals, organizations, or

institutions. These materials are carefully maintained to allow easy access, retrieval, and long-term preservation (Jeenapriyanka, n.d.). The National Archives of India (National Archives of India, n.d.), Abhilekh Patal (Abhilekh Patal, n.d.), Indian Culture Archives (Indian Culture, n.d.), Archives of Indian Labour (V. V. Giri National Labour Institute, n.d.), Archives of Contemporary India (Archives of Contemporary India,

n.d.), Prime Ministers Museum and Library Archive (Prime Ministers Museum and Library, n.d.), Archives at NCBS (Archives at NCBS, n.d.), Archive at IIT Madras (IIT Madras, n.d.), and Tamil Nadu Archives and Historical Research (Tamil Nadu Archives and Historical Research, n.d.) are the major archives in India. National Digital Library of India (NDLI) is a project sponsored by the Ministry of Education, Govt. of India. Developed by the Indian Institute of Technology Kharagpur, it is the most important initiative for the progress of the government's National Mission in Education through Information and Communication Technology (NMEICT). It is a single-window platform providing learning resources that make e-learning and education accessible to all. It brings digital repositories from India and the world to users of all demographics (National Digital Library of India Project, n.d.).

2. Related Works

In order to comprehend the National Digital Library of India's (NDL India) benefits, characteristics, and collection in the global digital arena, Bashir et al. (2019) made an effort to give an overview of the organization. The strategic importance of the NDLI Club program, which was started in 2021, in promoting education accessibility, especially in the Northeast, was examined by Mukherjee and Sutradhar (2024). Leveraging the framework laid out in the National Education Policy (NEP) 2020, NDLI Clubs serve as innovative platforms for activity-based learning, aligning with the educational objectives outlined in the policy. This paper delved into the spatial distribution and contextualization of NDLI Clubs, with a particular focus on the unique dynamics of Northeast India. Singh (2022) described the COVID-19 research repository assimilated by NDLI, and detailed investigative research has been made about various resources and materials available for the research community. This paper provides a structured overview of the COVID-19 research repository facilitated by NDLI to provide access to related research activities by scientists and researchers. Ashkar et al. (2022) explored the state of the art of NDLI among the Central University Kerala users and found the NDLI's resources and services are not fully exploited, despite the fact that the majority of students are aware of them. Many of the respondents are still ignorant of the NDLI's instruments and have

not used the services, which are crucial to their research and academic work. Wagh (2025) explained the key features, advantages, and primary services of NDLI, including its Institutional Digital Repository (IDR) hosting services, Digital Preservation Centre, and NDLI Club program. The analysis of 42 institutions partnering with NDLI to host their IDRs reveals a diverse mix of institutions, varying content quantities, and a growing trend of adoption over time.

3. Research Questions

- What is the distribution of digital cultural archive contents across different content providers in the National Digital Library of India?
- Which content providers contribute the highest and lowest number of cultural archive resources in NDLI?
- Which Indian languages are represented in the digital cultural archive contents of the National Digital Library of India (NDLI)?
- Which foreign languages contribute digital cultural archive contents to the NDLI repository?

4. Research Methodology

4.1. Search Strategy

For data collection, a systematic search strategy was adopted using the "Cultural Archives" category available on the National Digital Library of India (NDLI) portal. Relevant metadata, including the content provider name, document type, language, and the number of digital resources, were extracted and recorded for analysis.

4.2. Date of Data Extraction

The cultural archive data for this study were extracted from the NDLI portal on 20 December 2025.

4.3. Inclusion Criteria

The analysis only considered resources that were under the Cultural Archives category and were available via the NDLI site. 78 content providers provided archival resources for the investigation, which included 74 different document categories, 31 Indian languages, and 33 foreign languages.

4.4. Exclusion Criteria and Data Cleaning

During the data cleaning procedure, duplicate records, missing metadata information, and unrelated resources

were eliminated. Before statistical analysis, the gathered data were validated, arranged, categorized, and tabulated to guarantee data accuracy and consistency.

4.5. Data Analysis

Data collection, data cleaning, coding, classification, tabulation, and descriptive statistical analysis were all done using Microsoft Excel 2021. Microsoft Excel's built-in functions were used to calculate frequency counts, percentages, ranking, mean, median, standard deviation, and range. The processed dataset was imported into IBM SPSS Statistics (Version 29.0) for advanced statistical analysis in order to offer a more methodical analytical framework. To evaluate the relative variability in the distribution of archival resources, the Coefficient of Variation (CV) was computed. The concentration of digital resources across content providers was measured using the Herfindahl–Hirschman Index (HHI), and the degree of inequality in the distribution of archival resources was assessed using the Gini Coefficient. The study was able to quantify inequality, concentration, and variability using analytical metrics that went beyond basic descriptive statistics.

The statistical analyses were performed using the following equations:

- **Coefficient of Variation (CV):**

$$CV = \frac{SD}{Mean} \times 100$$

- **Herfindahl–Hirschman Index (HHI):**

$$HHI = \sum p_i^2$$

where (p_i) represents the percentage share of the i th content provider.

- **Gini Coefficient:**

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2 \bar{x}}$$

where (x_i) and (x_j) denote the number of resources contributed by individual content providers, (n) is the total number of providers, and $(\bar{x})$ is the mean number of resources.

Established statistical benchmarks were used to interpret the analytical data. Low concentration is indicated by HHI values below 1,500, moderate concentration is indicated by values between 1,500 and 2,500, and

high concentration is indicated by values above 2,500. A more equal distribution is indicated by values of the Gini Coefficient that are closer to 0, whereas a highly unequal distribution is indicated by values that are closer to 1. The relative dispersion of archival resources was assessed using the Coefficient of Variation, where larger values denote more variability in the distribution. A thorough assessment of the distribution, variability, concentration, and inequality of digital cultural archive resources across content providers, document types, and language-wise collections in the National Digital Library of India (NDLI) was made possible by the combination of descriptive statistics and analytical indices. Compared to descriptive statistics alone, this analytical framework offers a more thorough analysis of the distribution patterns.

5. Objectives of the Study

These are the major objectives of the study.

- To find out the distribution of digital content by content providers in the National Digital Library of India.
- To identify the types of documents.
- To examine the Indian language-wise distribution of digital content.
- To assess the foreign language-wise distribution of digital content.

6. Results and Discussion

6.1. Distribution of Digital Contents by Content Providers

Table 1: Distribution of Digital Contents by Content Providers

S. No.	Name of the Content Providers	No. of Contents	%	Rank
1	World eBook Library	2631449	33.29	1
2	Internet Archive	1618325	20.47	2
3	Library of Congress - Photographs	906724	11.47	3
4	HathiTrust Digital Library	870800	11.02	4
5	Digital Library of India	533459	6.75	5
6	Library of Congress - Manuscripts/Mixed Material	163642	2.07	6
7	Librivox	161455	2.04	7

8	Library of Congress - Books/Printed Material	153860	1.95	8
9	Library of Congress	148783	1.88	9
10	Art Institute of Chicago	121653	1.54	10
11	Indian Culture	109385	1.38	11
12	Getty: Museum Collection	89288	1.13	12
13	Tamilnaiyam Digital Library	75709	0.96	13
14	Library of Congress - Maps	54020	0.68	14
15	Project Gutenberg	51330	0.65	15
16	IGNCA - Archaeological Sites of India	27840	0.35	16
17	IGNCA - National Cultural Audiovisual Archives	21114	0.27	17
18	Digital Himalaya	19920	0.25	18
19	Internet Archive: Cultural Resources of India	19394	0.25	18
20	South Asia Archive	15692	0.2	20
21	Odia Bibhaba	15098	0.19	21
22	IGNCA - ASI Books	12558	0.16	22
23	The Art of Living	9929	0.13	23
24	Tribal Repository	9851	0.12	24
25	E-bharatisampat	8308	0.11	25
26	Faded Page	8157	0.1	26
27	Rare Book Society of India	4406	0.06	27
28	Rekhta	4391	0.06	27
29	IGNCA - Vedic Heritage Portal	3388	0.04	29
30	Sahapedia	3543	0.04	29
31	UNESCO Digital Library (UNESDOC)	2976	0.04	29
32	Culture and Archive - Assorted Archive	2259	0.03	32
33	Global Grey ebooks	2591	0.03	32
34	Prasar Bharati - DD Archives	2465	0.03	32
35	Society for Natural Language Technology Research	2176	0.03	32

36	The Heritage Lab	2294	0.03	32
37	Bharatiya Jnana Sarita	1851	0.02	37
38	D'source	1191	0.02	37
39	Gopa Bagchi Audio Books	1748	0.02	37
40	IGNCA - Content Development in Indian Language Network	1291	0.02	37
41	Tribal Digital Document Repository	1568	0.02	37
42	IGNCA - Devanarayana-An Oral Epic	445	0.01	42
43	IGNCA - Man and Mask	579	0.01	42
44	IGNCA - Sanskrit Text Repository	406	0.01	42
45	JAINpedia	864	0.01	42
46	Literature Collection	502	0.01	42
47	Odia Virtual Academy	416	0.01	42
48	Odisha State Tribal Museum	545	0.01	42
49	Presidency Alumni Association Calcutta	1134	0.01	42
50	Reading & Recreation - Assorted	700	0.01	42
51	Ambedkar Foundation	84	0	51
52	Ancient Engineering Technologies	21	0	51
53	Centre for Studies of Tradition and Systems (CSTS)	5	0	51
54	CSSSC Periodical Collection - Bengal Library Catalogue of Records	1	0	51
55	IGNCA - Arts & Crafts of North East	59	0	51
56	IGNCA - Begum Akhtar Portal	243	0	51
57	IGNCA - Brihadiswara Temple	20	0	51
58	IGNCA - Ganjifa Cards	135	0	51
59	IGNCA - Gita Govinda	270	0	51

60	IGNCA - Intangible Cultural Heritage of India	82	0	51
61	IGNCA - Kalyani Kosh	1	0	51
62	IGNCA - Manuscripts	65	0	51
63	IGNCA - South Asia Digital Images	42	0	51
64	IGNCA - Tribal Art & Culture	205	0	51
65	International Encyclopaedia of the First World War (1914-1918)	13	0	51
66	Mahakavya	251	0	51
67	National Mission for Manuscripts	296	0	51
68	Open Books: University of Edinburgh	342	0	51
69	Press Information Bureau: Kolkata	24	0	51
70	Ramakrishna Math and Ramakrishna Mission - Advaita Ashrama	119	0	51
71	Sahitya Akademi	53	0	51
72	Satyajit Ray Society	39	0	51
73	School of Cultural Texts and Records (SCTR), Jadavpur University	381	0	51
74	Shri Prannath Gyanpeeth	69	0	51
75	Sri Ma Trust	29	0	51
76	Tara Prakashana	2	0	51
77	Udbodhan: Ramakrishna Math, Baghbar	1	0	51
78	Valley of Words	133	0	51
	Total	7904457	100	

Table 1 presents the distribution of digital cultural archive resources contributed by 78 content providers in the National Digital Library of India (NDLI). The analysis reveals a highly concentrated pattern of content distribution, with a small number of major repositories contributing the majority of the available resources. The World eBook Library is the largest contributor

with 2,631,449 resources (33.29%), followed by the Internet Archive with 1,618,325 resources (20.47%), the Library of Congress – Photographs with 906,724 resources (11.47%), HathiTrust Digital Library with 870,800 resources (11.02%), and the Digital Library of India with 533,459 resources (6.75%). Collectively, these five repositories contribute approximately 83% of the total cultural archive resources available in NDLI, indicating a strong concentration of digital collections among a limited number of content providers. The distribution further demonstrates a substantial disparity between major and minor contributors. While hundreds of thousands or even millions of digital items can be found in the top-ranked repositories, many suppliers simply contribute a few hundred or fewer resources. This large variation points to an unequal pattern of resource aggregation in the NDLI ecosystem, where the collection is dominated by a small number of reputable repositories. Large-scale digital resources are made more accessible by this concentration, but it also highlights an imbalance in the representation of smaller institutional and speciality repositories. A comparison of Indian and foreign repositories reveals variations in their contributions as well. International repositories, including the World eBook Library, Internet Archive, Library of Congress, and HathiTrust Digital Library, occupy the leading positions and together account for a substantial proportion of the digital collections. Among the Indian repositories, the Digital Library of India ranks fifth with 533,459 resources, followed by Indian Culture with 109,385 resources and Tamilnaiyam Digital Library with 75,709 resources. These repositories play a significant role in preserving and disseminating India's cultural heritage, although their contributions remain considerably smaller than those of the largest international repositories. At the other end of the distribution, repositories such as the School of Cultural Texts and Records (SCTR), Jadavpur University (381 resources), Open Books: University of Edinburgh (342), National Mission for Manuscripts (296), IGNCA – Gita Govinda (270), Mahakavya (251), IGNCA – Begum Akhtar Portal (243), IGNCA – Tribal Art & Culture (205), and Valley of Words (133) contribute relatively limited collections. Despite their smaller size, these repositories provide specialised and culturally significant resources that complement the broader collections maintained by the larger repositories.

6.2. Archival Content Types

Table 2: Archival Content Types

S. No.	Document Type	Nos.	Rank
1	Books	5855583	1
2	Manuscripts	176509	2
3	Correspondence	112384	3
4	Visual Artwork	102313	4
5	Report	61563	5
6	Map	54395	6
7	Notes	50343	7
8	Journal	45065	8
9	Music Notation	43806	9
10	Article	41681	10
11	Painting	33902	11
12	Magazine	32467	12
13	Music	13055	13
14	Poetry	12861	14
15	Thesis	10177	15
16	Catalogue	8940	16
17	Media Article	7983	17
18	Video Lecture	4783	18
19	Annual Report	3652	19
20	Monograph	3623	20
21	Collection	3467	21
22	Proceeding	3458	22
23	Diary	2,902	23
24	Gazetteer	2735	24
25	Biography	2626	25
26	Story	2535	26
27	Manual	2,533	27
28	Broadcast	2471	28
29	Summary	1933	29
30	Autobiography	1628	30
31	Newspaper Article	1621	31
32	Review	1331	32
33	Conference Proceedings	1326	33
34	Technical Report	1291	34
35	Transcript	1272	35

36	Historical Record	943	36
37	Series	925	37
38	Interview	874	38
39	Play	774	39
40	Survey	590	40
41	Film	515	41
42	Almanac	478	42
43	Dataset	440	43
44	Letter	414	44
45	Directory	379	45
46	Issue	283	46
47	Law Act	275	47
48	Presentation	251	48
49	Bibliography	222	49
50	Notice	205	50
51	Chapter	202	51
52	Discussion	179	52
53	News Letter	154	53
54	Law Report	144	54
55	Synopsis	140	55
56	Case Study	133	56
57	Audio Lecture	129	57
58	Leaflet	81	58
59	Appendix	75	59
60	Brochure	68	60
61	Educational Game	47	61
62	Podcast Presentation	40	62
63	Plan	28	63
64	Project Report	28	63
65	Activity	23	65
66	Calendar	15	66
67	Quiz	14	67
68	Simulation	13	68
69	Book Review	11	69
70	Question	7	70
71	Lecture Notes	6	71
72	Training Manual	6	71
73	Solution	5	73
74	Album	2	74

Table 2 presents the distribution of archival content types available in the National Digital Library of India (NDLI). The analysis identified 74 distinct archival content types, reflecting the multidisciplinary nature of the repository. However, the distribution of these content types is highly uneven, with a small number of document categories accounting for the majority of the digital archive. Books constitute the largest collection with 5,855,583 resources, followed by manuscripts (176,509), correspondence (112,384), visual artwork (102,313), reports (61,563), maps (54,395), notes (50,343), journals (45,065), music notation (43,806), and articles (41,681). The overwhelming dominance of books demonstrates that NDLI primarily functions as a text-oriented digital repository, while other archival formats contribute comparatively smaller collections. A comparative analysis of the document categories indicates that traditional documentary resources, including books, manuscripts, correspondence, reports, journals, and maps, substantially outnumber multimedia and interactive archival formats. This distribution reflects the historical emphasis of digitisation initiatives on preserving textual and documentary heritage, which are comparatively easier to digitise, catalogue, and disseminate through digital library platforms. The significant representation of manuscripts and correspondence also highlights NDLI's contribution to preserving valuable primary historical sources and documentary heritage. The repository also contains a moderate volume of specialised archival resources, including paintings (33,902), magazines (32,467), music (13,055), poetry (12,861), and theses (10,177). These collections enhance the disciplinary diversity of NDLI by supporting research in literature, fine arts, music, history, and cultural studies. Nevertheless, their representation remains considerably lower than that of conventional textual resources, indicating an imbalance in the composition of the digital collections. In contrast, emerging and non-traditional digital content types remain underrepresented. Resources such as historical records (943), interviews (874), films (515), datasets (440), educational games (47), podcast presentations (40), simulations (13), lecture notes (6), training manuals (6), solutions (5), and albums (2) constitute only a small fraction of the overall repository. The comparatively limited

availability of these multimedia, interactive, and research data resources suggests that NDLI has yet to fully diversify its archival collections in line with evolving digital scholarship and contemporary digital preservation practices.

6.3. Indian Language-wise Distribution of Digital Contents

Table 3: Indian Language Wise Distribution of Digital Contents

S. No.	Language Name	Contents	%	Rank
1	Hindi	66025	17.3	1
2	Sanskrit	63629	16.67	2
3	Tamil	56239	14.73	3
4	Gujarati	44075	11.55	4
5	Bengali	43461	11.39	5
6	Urdu	38479	10.08	6
7	Telugu	31661	8.29	7
8	Odia/ Oriya	15411	4.04	8
9	Marathi	5465	1.43	9
10	Kannada	4970	1.3	10
11	Manipuri	2928	0.77	11
12	Assamese	2227	0.58	12
13	Rajasthani	1830	0.48	13
14	Garo	1117	0.29	14
15	Malayalam	1058	0.28	15
16	Nepali	1021	0.27	16
17	Khasi	605	0.16	17
18	Panjabi/Punjabi	588	0.15	18
19	Kashmiri	505	0.13	19
20	Sindhi	163	0.04	20
21	Chhattisgarhi	126	0.03	21
22	Konkani Sanskrit	30	0.01	22
23	Bhojpuri	26	0.01	23
24	Bodo/Boro	2	0	24
25	Bundeli/Bundel Khandi	9	0	24
26	Dogri	3	0	24
27	Garhwali	15	0	24
28	Maithili	11	0	24
29	Munda	2	0	24
30	Santali	11	0	24
31	Sikkimese	2	0	24

Table 3 presents the Indian language-wise distribution of digital archival resources available in the National Digital Library of India (NDLI). The analysis identified archival resources in 31 Indian languages, demonstrating NDLI's commitment to preserving India's linguistic and cultural diversity. However, the distribution of resources is highly uneven, with a small number of languages accounting for the majority of the available collections. Hindi contributes the largest share with 66,025 resources (17.30%), followed closely by Sanskrit (63,629; 16.67%) and Tamil (56,239; 14.73%). Gujarati (44,075; 11.55%), Bengali (43,461; 11.39%), Urdu (38,479; 10.08%), Telugu (31,661; 8.29%), and Odia/Oriya (15,411; 4.04%) also represent substantial proportions of the repository. Collectively, these eight languages account for the overwhelming majority of Indian language resources, indicating a strong concentration of digital archival collections in a limited number of widely spoken and historically significant languages.

The representation of Indian languages in the repository varies substantially, according to a comparative investigation. Compared to many other regional languages, the collections of classical and regionally dominating languages such as Sanskrit, Tamil, Bengali, Gujarati, Telugu, and Urdu are far larger. This pattern may reflect differences in the availability of digitised heritage materials, historical publishing activity, institutional digitisation initiatives, and the level of support received through national and state-level cultural preservation programmes. The strong representation of these languages enhances access to India's literary, historical, and cultural heritage while supporting research across multiple disciplines. Languages such as Marathi (5,465), Kannada (4,970), Manipuri (2,928), Assamese (2,227), Rajasthani (1,830), Garo (1,117), Malayalam (1,058), and Nepali (1,021) contribute moderate numbers of archival resources. These collections demonstrate ongoing efforts to preserve regional linguistic heritage; however, their representation remains considerably lower than that of the leading languages, highlighting disparities in the distribution of digital collections. At the lower end of the distribution, languages including Khasi (605), Punjabi (588), Kashmiri (505), Sindhi (163), Chhattisgarhi (126), Konkani (30), Bhojpuri (26), Garhwali (15), Maithili (11), Santali (11), Bundeli/Bundel Khandi (9), Dogri (3), and Bodo/

Boro, Munda, and Sikkimese (2 each) contribute only a limited number of archival resources. The comparatively low representation of these languages suggests that several regional and indigenous linguistic communities remain underrepresented in the NDLI repository. This imbalance may affect the visibility and long-term preservation of valuable cultural knowledge associated with these communities.

6.4. Foreign Language-wise Distribution of Digital Contents

Table 4: Foreign Language-wise Distribution of Digital Contents

S. No.	Language Name	Nos.	%	Rank
1	English	6321774	89.39	1
2	French	208893	2.95	2
3	German	182647	2.58	3
4	Chinese	80977	1.14	4
5	Spanish	80383	1.14	4
6	Italian	72096	1.02	6
7	Portuguese	17659	0.25	7
8	Dutch	15130	0.21	8
9	Persian	14748	0.21	8
10	Russian	15094	0.21	8
11	Arabic	13791	0.19	11
12	Japanese	7222	0.1	12
13	Danish	5655	0.08	13
14	Swedish	5959	0.08	13
15	Greek	4890	0.07	15
16	Czech	3639	0.05	16
17	Polish	3685	0.05	16
18	Hungarian	3138	0.04	18
19	Catalan	2456	0.03	19
20	Finnish	2115	0.03	19
21	Hebrew	2468	0.03	19
22	Norwegian	1345	0.02	22
23	Ukrainian	1511	0.02	22
24	Bulgarian	390	0.01	24
25	Croatian	484	0.01	24
26	Indonesian	408	0.01	24
27	Irish	576	0.01	24
28	Korean	945	0.01	24
29	Romanian	995	0.01	24

30	Turkish	759	0.01	24
31	Afrikaans	250	0	31
32	Malay	138	0	31
33	Slovak	273	0	31

Table 4 presents the foreign language-wise distribution of digital archival resources available in the National Digital Library of India (NDLI). The analysis identified archival resources in 33 foreign languages, indicating the multilingual and international scope of the repository. However, the distribution of resources is highly concentrated, with English overwhelmingly dominating the foreign language collections. 6,321,774 resources (89.39%) are in English, followed by 208,893 resources (2.95%) in French and 182,647 resources (2.58%) in German. Italian accounts for about 1.02% of the collections, while Chinese and Spanish both contribute roughly 1.14%. These six languages together make up the great bulk of NDLI's foreign language resources, indicating a significant concentration of digital collections in a small number of globally recognised languages. Significant differences exist in foreign language collections, according to a comparative analysis. English is the primary language of worldwide scholarly communication, digital publishing, and international cultural heritage repositories, which is reflected in its overwhelming domination. European digital libraries and cultural institutions have made a substantial contribution to the NDLI collection, as seen by the moderate representation of European languages including French, German, Spanish, Italian, Portuguese, Dutch, Russian, and Swedish. These collections facilitate comparative and multidisciplinary research by expanding access to global literature, historical records, and cultural resources. Despite their worldwide cultural and academic significance, Asian and Middle Eastern languages—such as Chinese, Japanese, Persian, Arabic, and Korean—have relatively little representation. This distribution suggests that the availability of foreign language resources in NDLI is influenced by existing international digitisation initiatives and repository collaborations rather than reflecting an equal representation of global linguistic diversity. Consequently, some major world languages remain comparatively underrepresented within the repository. Languages such as Ukrainian (1,511), Norwegian (1,345), Romanian (995), Korean (945),

Turkish (759), Irish (576), Croatian (484), Indonesian (408), Bulgarian (390), Slovak (273), Afrikaans (250), and Malay (138) contribute relatively small collections. Although these resources enhance the multilingual character of NDLI, their limited representation indicates opportunities to strengthen partnerships with international libraries, archives, and cultural heritage organisations to improve the diversity and comprehensiveness of foreign language collections.

6.5. Analytical Measures of Distribution of Digital Contents by Content Providers

Table 5: Analytical Measures of Distribution of Digital Contents by Content Providers

Statistical / Analytical Measure	Value
Mean	101,339.19
Median	2,408
Standard Deviation	373,735.76
Coefficient of Variation (CV)	368.80%
Range	2,631,448
Herfindahl–Hirschman Index (HHI)	1,849.58
Gini Coefficient	0.914

- **Coefficient of Variation (CV):**

$$CV = \frac{373,735.76}{101,339.19} \times 100$$

$$CV = 3.6880 \times 100$$

$$CV = 368.80\%$$

Coefficient of Variation = 368.80%

- **Herfindahl–Hirschman Index (HHI):**

$$p_i = \frac{x_i}{\sum x} \times 100$$

Table 6: Herfindahl–Hirschman Index (HHI) of Digital Cultural Archive Resources by Content Providers

Provider	Share (%)	Share ²
World eBook Library	33.29	1108.22
Internet Archive	20.47	419.02
Library of Congress–Photographs	11.47	131.56

HathiTrust Digital Library	11.02	121.44
Digital Library of India	6.75	45.56
Remaining 73 providers	—	23.78

Therefore,

$$HHI = 1108.22 + 419.02 + 131.56 + 121.44 + 45.56 + \dots$$

$$HHI = 1849.58$$

• **Gini Coefficient:**

$$n = 78$$

$$\bar{x} = 101,339.19$$

The calculation compares every provider with every other provider, computes the absolute differences, sums them, and divides by $2n^2\bar{x}$.

Using the complete dataset,

$$G = 0.914$$

Table 5 presents the analytical measures describing the distribution of digital contents contributed by the 78 content providers in the National Digital Library of India (NDLI). More than half of the content providers offer very small collections, as evidenced by the mean contribution of 101,339.19 resources per provider, while the median is only 2,408. A highly skewed distribution with a small number of repositories contributing extraordinarily huge amounts of digital resources is suggested by the significant discrepancy between the mean and median. The coefficient of variation (CV = 368.80%) and standard deviation (373,735.76) show incredibly high diversity among content sources. The distribution is substantially dispersed rather than uniform since the CV significantly surpasses 100%, indicating that contributions differ significantly amongst repositories. The distribution of digital resources among content providers is considerably concentrated, according to the Herfindahl–Hirschman Index (HHI = 1,849.58). This result implies that although a large number of repositories take part in NDLI, a significant amount of the collection is held by a comparatively small number of important repositories. This concentration pattern is greatly influenced by the prominence of repositories like the World eBook Library, Internet Archive, Library of Congress, and HathiTrust Digital Library. An extremely high degree of inequality in the allocation of digital resources is further supported by the Gini index (0.914). A Gini value near one means

that the majority of content providers provide relatively tiny collections, and digital content is concentrated among a small number of repositories. This analytical conclusion shows that the NDLI repository is marked by a very unequal contribution structure and statistically validates the observed distribution pattern. All things considered, these analytical metrics offer a methodical assessment of the distribution of digital resources that goes beyond descriptive data. The results show that NDLI has significant inequality in content contributions, moderate concentration, and high variability. These findings indicate that in order to create a more representative and balanced national digital cultural heritage collection, smaller institutional, regional, and specialized repositories should be encouraged to participate more.

7. Discussion

The present study presents a detailed analysis of the distribution of cultural archival resources available in the National Digital Library of India (NDLI) by assessing content providers, archival content types, and language-wise collections. A methodical evaluation of the variability, concentration, and inequality of digital cultural archive materials is possible through the use of descriptive statistics as well as the Coefficient of Variation (CV), Herfindahl–Hirschman Index (HHI), and Gini Coefficient. The NDLI repository shows a somewhat concentrated but extremely uneven distribution of digital resources, according to the content providers' study. While the Gini Coefficient (0.914) shows significant disparity in content contributions, the computed HHI (1,849.58) shows a modest amount of concentration. Significant variation amongst repositories is further confirmed by the extremely high Coefficient of Variation (368.80%). The Internet Archive (20.47%), Library of Congress – Photographs (11.47%), HathiTrust Digital Library (11.02%), and the World eBook Library alone account for 33.29% of all cultural archive materials. These repositories collectively provide more than three-fourths of the NDLI's total archival resources, indicating the repository's heavy reliance on a comparatively small number of sizable international digital libraries. The Digital Library of India, Indian Culture, Tamilnaiyam Digital Library, and several IGNCA repositories are among the Indian contributors

who contribute significantly to the preservation of indigenous knowledge, local history, and cultural legacy. However, the relatively modest contributions from many institutional repositories point to the necessity of promoting wider involvement in order to create a more equitable national digital archive.

The NDLI collection is primarily made up of traditional textual resources, according to the examination of archival content types. The vast majority of the collections consist of books, which are followed by manuscripts, correspondence, visual artwork, reports, maps, journals, and music notation. The historical emphasis of digitisation programs on documentary and literary heritage preservation is reflected in this distribution. On the other hand, only a small percentage of the repository is made up of multimedia and developing digital resource categories, such as databases, podcasts, instructional games, simulations, and audio lectures. This disparity implies that in order to promote modern research, digital humanities, and interdisciplinary scholarship, future digitisation efforts should concentrate on increasing born-digital and multimedia heritage resources.

An uneven distribution of archival resources among Indian languages is further shown by the language-wise analysis. The bulk of Indian language collections are in Hindi, Sanskrit, Tamil, Gujarati, Bengali, Urdu, and Telugu, suggesting that these languages have profited from more robust digitisation efforts and increased access to historical materials. While Tamil, Bengali, Gujarati, Telugu, and Urdu collections show active regional attempts to preserve linguistic and cultural history, the significant representation of Sanskrit emphasises the preservation of ancient Indian knowledge systems, manuscripts, and traditional literature. Languages like Bodo, Santali, Dogri, Maithili, Munda, and Sikkimese, on the other hand, are still hardly represented, demonstrating differences in the digitisation of indigenous and regional language resources. A more inclusive preservation of India's multilingual cultural legacy will result from improving the representation of these under-represented languages.

English predominates in the repository, making up 89.39% of all foreign language materials, according to the examination of foreign language collections. The global predominance of English in scholarly communication, digital publishing, and international

digital library projects is reflected in this study. Asian, African, and other global languages are somewhat under-represented, while European languages, such as French, German, Spanish, and Italian, make up the second-largest category of collections. Languages like Afrikaans, Malay, Slovak, Bulgarian, Croatian, and Indonesian are comparatively few, which raises the possibility of increasing international cooperation with international cultural heritage organisations to enhance multilingual representation within NDLI.

8. Conclusion

Archival resources play a fundamental role in preserving historical records, cultural heritage, and indigenous knowledge, thereby supporting education, research, and lifelong learning. The National Digital Library of India's (NDLI) cultural archive collections were methodically examined in this study by looking at the distribution of content providers, archival material types, and language-wise resources. To assess the variability, concentration, and inequality in the distribution of digital cultural archive resources, the study used descriptive statistics along with the Coefficient of Variation (CV), Herfindahl–Hirschman Index (HHI), and Gini Coefficient.

The results show that the cultural archive collections in the NDLI repository show significant disparity, moderate concentration, and high variability. While Indian repositories are crucial to the preservation of indigenous knowledge and regional traditions, a comparatively limited number of important foreign repositories contribute the majority of the available resources. Multimedia and new digital archive forms are still somewhat under-represented in the collection, which is primarily made up of traditional textual items, especially books, manuscripts, reports, and journals. In a similar vein, the language-wise study shows that while a number of regional and minority languages have very little representation, a few major Indian and foreign languages make up the majority of the collections.

These results show that even though NDLI has effectively established itself as a national platform for granting access to digital cultural material, there are still chances to enhance the balance and diversity of its holdings. The comprehensiveness, inclusivity, and long-term sustainability of the repository would be further improved by strengthening the involvement of

regional institutions, growing multilingual collections, and increasing the digitisation of multimedia and born-digital resources. By offering a methodical analytical evaluation of NDLI's cultural archive holdings and evidence that might assist future collection development policies and digitisation strategies, the study advances the field of digital library and cultural heritage research.

9. Recommendations

- Educational institutions and librarians should upload their archival contents to the NDLI portal regularly to improve the availability of cultural archive resources.
- Greater efforts should be taken to preserve and digitize regional, tribal, and underrepresented language resources in the NDLI repository.
- NDLI and participating institutions should encourage the inclusion of multimedia and emerging digital archival formats such as datasets, podcasts, simulations, audio lectures, and educational games.
- More Indian universities, libraries, museums, and cultural organizations should participate in contributing archival resources to strengthen indigenous knowledge preservation.
- Awareness programs and user orientation initiatives should be conducted among students, researchers, and faculty members to promote the effective use of NDLI cultural archive resources for teaching, learning, and research activities.

Acknowledgement

The author acknowledges the National Digital Library of India (NDLI) portal for providing access to publication data used in this study. The author also acknowledges that the NDLI portal is developed and maintained by IIT Kharagpur.

Authorship Contribution

The author was responsible for all aspects of the study, including conceptualization, methodology, data collection and analysis, and preparation of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The author declares no conflict of interest related to this study.

Declarations

The author declares that this work is original and has not been submitted elsewhere for publication. All data, methodologies, and system components have been developed and reported in accordance with accepted academic standards. All referenced materials have been duly cited, and the author accepts full responsibility for the integrity and accuracy of the findings presented.

Ethical Approval

Ethical approval was not required for the study.

Figure Permissions

The author declares that all figures and tables presented in this work are original and have not been previously published or are otherwise used in accordance with applicable permissions and copyright requirements.

Data Availability Statement

The data supporting the findings of this study are openly available through the National Digital Library of India (NDLI). The dataset can be accessed through the official website: <https://www.ndli.gov.in/>.

Human and Animal Rights

No human participants or animals were involved in this research.

AI Disclosure Statement

No artificial intelligence (AI) tools were used in the preparation of this manuscript.

References

- Abhilekh Patal. (n.d.). *Abhilekh Patal*.
<https://www.abhilekh-patal.in/>
- Archives at NCBS. (n.d.). *Home*.
<https://archives.ncbs.res.in/>
- Archives of Contemporary India. (n.d.). *Archives of Contemporary India*. Ashoka University.
<https://archives.ashoka.edu.in/>

- Ashkar, K., Sony, S. R., & Arun, V. R. (2022). Contribution of National Digital Library of India on the paradigm shift in research and education: A study based on the Central University Kerala. *Library Philosophy and Practice (e-journal)*, Article 7346. <https://digitalcommons.unl.edu/libphilprac/7346/>
- Bashir, B., Nasreen, N., & Loan, F. A. (2019). National Digital Library of India: An overview. *Library Philosophy and Practice (e-journal)*, Article 2601. <https://digitalcommons.unl.edu/libphilprac/2601/>
- IIT Madras. (n.d.). *Archive of IIT Madras*. <https://archive.iitm.ac.in/>
- Indian Culture. (n.d.). *Archives*. <https://indianculture.gov.in/archives>
- Jeenapriyanka, J. (n.d.). *Basics of museum & archive*. Fakir Mohan University. https://fmuniversity.nic.in/getdata?dir=deptnotice&rid=notice2004_29520251748500820482.pdf
- Mukherjee, A., & Sutradhar, B. (2024). Catalysing Change: A Case Study of NDLI Club as Outreach Model for Education Access in NE India. <https://ir.inflibnet.ac.in/items/92608bcd-0463-40b1-ac52-4cf3c65debad>
- National Archives of India. (n.d.). *Official website of the National Archives of India*. <https://nationalarchives.nic.in/>
- National Digital Library of India Project. (n.d.). *About us*. <https://project.ndl.gov.in/about-us/>
- Prime Ministers Museum and Library. (n.d.). *Prime Ministers Museum and Library*. <https://pmml.nic.in/>
- Singh, S. (2022). Role of National Digital Library of India (NDLI) for facilitating open access resources (OARs): An investigation on COVID-19 research repository. *Digital Library Perspectives*, 38(4), 493–507. <https://doi.org/10.1108/DLP-08-2021-0072>
- Tamil Nadu Archives and Historical Research. (n.d.). *Tamil Nadu Archives and Historical Research*. Government of Tamil Nadu. <https://tamilnaduarchives.tn.gov.in/WEB/EN/>
- V. V. Giri National Labour Institute. (n.d.). *Digital archives*. <https://www.vvgnli.gov.in/en/archive/digital-archives>
- Wagh, R. D. (2025). NDLI: A catalyst for educational advancement through digital repositories, preservation, and community engagement. *International Journal of Innovative Science and Research Technology*, 10(9), 2380–2385. <https://doi.org/10.38124/IJISRT/25SEP1234>