

# Lidas: Content Development, Experiences and Challenges

Mehaboobullah, K  
Reference Assistant, Kerala University Library  
Palayam, Thiruvananthapuram  
[mehaboobkc10@gmail.com](mailto:mehaboobkc10@gmail.com)



**ABSTRACT:** *The present study intends to analyse the digital archive of the Kerala University Library (LIDAS). The purpose of this paper is to analyse the documents selected for digitisation, content development strategies, digitisation process and challenges faced during the process. The study applied Literature survey method and analysing different records of the digitisation section of the Kerala University Library. Relevant literature related to digitisation and its implications has been reviewed. The study examine the process of the digitisation, planning for digitization, software details, challenges in the project, language-wise distribution of documents, subject-wise distribution of digitized documents and year-wise distribution of digitized documents. Lack of proper digitization policy at the University level is one of the most important problems for starting the digitisation of old and rare documents. In order to solve this problem, UGC have to come forward to formulate a common policy for digitisation and coordinating the digitisation activities taking place at various Universities in India.*

**Keywords:** Digital Archive, Digital Library, LIDAS, Kerala University Library, Digital Preservation

**Received:** 19 October 2018, Revised 13 November 2018, Accepted 29 November 2018

**DOI:** 10.6025/jio/2019/9/1/1-9

© 2019 DLINE. All Rights Reserved

## 1. Introduction

The recent development in IT has changed the concept of library and one of the most important developments in the library field is digitisation of library materials. Digitised resources have now become an important part of library resources. User community is getting more advantages through the digital resources better than conventional method of information system. Digital archiving and its allied concepts has brought a phenomenal change in the collection, preservation and dissemination of information. Kerala University Library, the mother University Library in Kerala which has been accumulated voluminous information over the years and most of them are very old, rare and valuable. These types of materials required special attention and digitisation is the best solution for keeping and preserving these materials for the longer period of time. Hence there was an urgent need to have a sound storage system and effective retrieval mechanism to deal with vast bits of information available in

the library. In this context, the Library has started its digitisation initiative (LIDAS-Library Digitisation and Archival System) of old and rare documents in the year 2009. The archival development has also imposed certain responsibilities and challenges to the professionals of the Kerala University Library.

## **2. Digital Library V/S Digital Archive**

We use digital Library to describe the situation where documents are stored in digital form. A digital Library is a library that contains materials in digitised form or access to internal or external resources that can be manipulated and delivered in many ways. But what of the older documents on paper with immense historical values that occupy so many spaces of the library. The solution is digitisation of rare and old materials of the existing collection for the purpose of preservation and posterity. Digital archives are distinct from digital libraries in the sense that digital libraries are repositories that collect and provide access to digital information, but may or may not provide for the long terms storage and access of that information.

## **3. Digitisation of Old and Rare Documents in the Kerala University Library**

Kerala University Library is the oldest and largest University Library in Kerala which was started in the year 1942. The library has more than 3.9 lakhs of books of which many are rare and old one. It has a comprehensive and specialised collection of rare and old documents even from 18th century onwards, which are not available elsewhere. Rare documents such as '*Krishnyanikalariyendumsamkshepedartham*'; the first printed book in Malayalam published in 1772; 'Kasiyatrappotta', an old travelogue (1872), Varthamanapusthakam vol.1(1936); the Census of Travncore 1891-1951; acts and proclamation of Travancore; Report on the administration of Travancore from 1866 onwards etc. are available in the section. The collection also includes some of the old Malayalam and English journals of the 19th and early 20th centuries. Most of these rare documents have been digitised and kept in the digital archives which comprises of 700000 of pages constituting approximately 1300 documents.

The Kerala University Library subscribes 7 English newspapers, 10 Malayalam newspapers and 3 Tamil newspapers and getting 6 papers, one English and 5 Malayalam newspapers as gifts constituting a total of 26 newspapers. Among these, 5 newspapers namely The Hindu, The New Indian Express (formerly The Indian Express), Malayala Manorama, Mathrubhumi and Kerala Kaumudi are bound periodically and kept in a separate room in the maintenance section. The old newspapers have large number of users since they contain various historical facts and which are of immense value to the research community. Moreover, the bound volumes of newspapers occupy a lion's share of the library's space which is affecting the stacking facility. It is not possible to digitise all the newspapers available in the Kerala University Library. Hence as a pioneer attempt, in the first phase, the library has digitised the newspaper 'Malayalarajyam', one of the old newspapers and the publishers have ceased its circulation since mid-seventeen and closed down. The rare documents are scanned, edited, indexed and uploaded in the archive using the LIDAS (Library Digitisation and Archival System) software which could be effectively retrieved by using the keywords contained in the documents. Separate terminals are provided to the users to search the digital archive. The archive is to be intended for 'fair use' by scholars of the library and accessibility is restricted within the library building.

The phase 1 project was outsourced to the C-DIT (Centre for Development of Imaging Technology) and work was completed in 2010. The library spent approximately 50 lakhs for this project. The digital archive is now widely used by the scholar not only from Kerala but from various other states and even from abroad. With the success of the Phase 1 of the digitisation, the Government of Kerala again allotted an amount of 50 lakhs specifically for the project of digitisation of old and rare document phase 2 and the work was started on 5th October 2016, by M/S Informatics Ltd, Bangalore.

## **4. Review of Literature**

The increasing interest in digitization is clearly apparent in the number of studies conducted recently. Bultmann and Hardy (2006) analyzed digital content in the UK Research Library and Archives sector. Sudhier (2007) explained the techniques of digital archives of manuscripts in the Manuscript Library, University of Kerala. Natarajan and Bhakta (2008) presented the issues and problems in building an digitization framework in India. The study listed the checklist for digitization with the key issues for managing the technology including the project planning, resource management, feasibility studies and costing of the project. Rafiq (2014) in his study assessed the opinions of the Pakistani library experts in terms of: organization for digitization work (execution of digitization work & staffing patterns);

funding patterns; and access patterns for digitization projects. The study opted qualitative research design and indepth interviews were conducted to collect the data from purposely selected library experts from different parts of Pakistan. The findings bring forward an outline of digitization framework in terms of execution of digitization work, staffing patterns, funding patterns and access patterns. Ringeland Ribak (2015) in their article offer some analytical comments on the digitization of the National Library of Israel (NLI), which has been converting materials into digital formats since 2007. This research draws upon Actor-Network Theory and is based on participant observation within the Digitization Center of the NLI. Mills (2015) conducted a study on User Impact on Selection, Digitization, and the development of Digital Special Collections. This article examines user impact on selection and the process of developing useful and usable digital special collections that balance user and institutional interests while satisfying the needs of key stakeholders. San and Peng (2017) studied the impact of copyright law on the digitization of library collections in academic libraries in Malaysia.

## **5. Objectives of the Study**

The main objectives of the study are:

- 1) To explain and analyse the criteria adopted for the selection of materials for digitisation and its process.
- 2) To discuss various challenges faced by the Library while implementing its digital archive.
- 3) To analyse the document-wise distribution of digitised materials of the Kerala University Library.
- 4) To find out the llanguage-wise distribution of digitised materials of the Kerala University Library.
- 5) To assess the subject-wise distribution of digitised materials of the Kerala University Library.
- 6) To examine the year-wise distribution of digitised materials of the Kerala University Library.

## **6. Methodology**

Primary data regarding the digitisation project of the Kerala University Library is collected from various records kept in the digitisation section of the Kerala University Library and from the digital archive software, LIDAS (Library Digitisation and Archival System).

## **7. Planning for Digitization**

Before starting the digitisation project, the Kerala University Library adopted the following steps for its successful completion.

### **7.1 Constituted an Expert Team**

An expert committee has been constituted by the library towards the implementation of the digital archive. The Committee comprises of the experts from the field of Information Technology, Library and Information Science and Malayalam language.

### **7.2 Formulated Criteria for Selection of Documents:**

The main criteria of selecting documents for the project were:

- 1) Rarity & Old
- 2) Historical importance
- 3) Copyright free and
- 4) Copyright owned

### **7.3 Task Force within the Library**

A task force comprising of the professionals from the different sections of the library has been constituted for smooth functioning of the project.

### **7.4 Hardware-software Evaluation and Selection**

Software and hardware components of the digital archive has been finalised before the actual implementation of the

project. Comparison of various digital library software, capturing devices, standards of formats etc where conducted.

### 7.5 Solution for Malayalam Inputting and Transliteration

Metadata of documents in regional languages and its keywords had to be transliterated in to English for easy retrieval. A standard transliteration method was developed for Malayalam sounds and letters under the guidance of the language expert in the committee.

## 8. Process of the Digitisation

The overall works of the Digitisation and Archival of the Kerala University Library can be broadly divided in to two:

### 8.1 Manual Works

- Identification of documents
- Manual recoding and cataloguing
- Metadata creation
- Identification of keywords

### 8.2 Computer Oriented Works

- Scanning/Capturing
- Editing (using photo editing software)
- Conversion in to PDF
- Metadata and keywords inputting

## 9. Software Details

Details regarding the software used for the purpose of digitisation in the Kerala University Library is given in the Table.1

S.I No	Name of the software	Font end	Backend	Company
1	LIDAS	PHP	MySQL	CDIT (with the support of the KUL)

Table 1. Software Details of the Project

The second phase of the digitisation was started on October 2016 by M/S Informatics Ltd, Bangalore which use DSPACE as Library digitization software.

## 10. Challenges in the Project

The main challenges of the project are:

### 10.1 Criteria Formulation for Selection of Documents

There are several approaches available to build digital archive such as retrospective conversion of collection (starting at A and ending up Z), digitisation of particular special collection, digitisation of high use materials in the library, Adhoc approach (where one digitised and stores materials as they are requested) etc. The committee examined the pros and cons of the above approaches and decided to select documents based on the four criteria ie. rarity, historical importance, copyright free and copyright owned materials. Copyright free documents are usually government documents (mainly documents related to Travancore government) or old books whose copyright is either expired or not applied. The library has digitised the copy right owned materials of the University like in-house reports, in-house publications, conference proceeding, hand books etc.

### 10.2 Software Development

This is probably the most important and central aspect of the technology infrastructure for the digital archive. The library evaluated some of the digital library experiments of that times and has reached a conclusion that no digital library of that time could replicate to the digital archive of the Kerala University Library since almost half of the of the documents were in the Malayalam and other regional languages. More over most of the documents were dealing with multiple subject contents which necessitated comprehensive metadata creation and keyword identification. Based on the inputs from the expert committee, the library entrusted CDIT, an agency under the Government of Kerala, to develop suitable software for the digital archive.

### 10.3 Problem of OCRing in Regional Language

OCR software is used for conversion of scanned text to machine readable text that can be searched or indexed. But it was not successful in the case of scanned regional language documents and thus necessitated page wise keyword identification and uploading same to the software for its easy and exhaustive retrieval. A standard transliteration method was also developed for Malayalam sounds and letters under the guidance of the language expert in the committee.

### 10.4 Copyright Issues

Copyright was one of the most serious barriers to digital archive development of the Kerala University Library. Copyright protects every form of expression in the literary and artistic domain. *It provides a bundle of exclusive economic rights to the owner to reproduce the work in any material form, to issue copies of the work to the public, to perform the work in public, to communicate the work to the public, to make any cinematograph film or sound recording in respect of the work, to make any translation of the work to make any adaptation of the work.* Library committee and digitization team conducted a series of enquiries and discussions about various aspects of copy right law before the actual commencement of the project and some of the important points are given below.

1) Infringement occurs when someone without the permission of the copyright owner does any of the exclusive acts, which copyright owner alone can do. Thus, scanning or photocopying a copyrighted book normally will amount to copyright infringement as the exclusive right to make copies of that book belongs to the copyright.

2) *Fair use provision in the copy right law helps the libraries for getting legal security towards the reproduction of its materials under certain circumstances. Fair uses are situations where unlicensed/unpermitted use of copy righted materials does not constitute infringement.* There are instances of use of a copyrighted work without copyright owner's consent of which the law takes no account as the law considers such uses to be fair and thus, not constituting infringement. Fair use is, thus, constitute a kind of limitation to the copyright owner's exclusive rights. This is recognized by international treaties and domestic laws.

The purpose of copyright is to strike a delicate balance between the author's economic rights and the right of the public at large. Fair use is one the tools that helps to maintain this delicate balance. *Fair Dealing of Library Exceptions in India [Section 52(1) of the Copyright Act, 1957] has the following subsections.*

(o) the making of not more than three copies of a book (including a pamphlet, sheet of music, map, chart or plan) by or under the direction of the person in charge of a non-commercial public library for the use of the library if such book is not available for sale in India; (<http://copyright.gov.in>)

(p) the reproduction, for the purpose of research or private study or with a view to publication, of an unpublished literary, dramatic or musical work kept in a library, museum or other institution to which the public has access:

Provided that where the identity of the author of any such work or, in the case of a work of joint authorship, of any of the authors is known to the library, museum or other institution, as the case may be, the provisions of this clause shall apply only if such reproduction is made at a time more than sixty years from the date of the death of the author or, in the case of a work of joint authorship, from the death of the author whose identity is known or, if the identity of more authors than one is known from the death of such of those authors who dies last. (<http://copyright.gov.in>)

Since the project was started in the year of 2009, only the above two sub clauses were helpful to reproducing library materials and which was ambiguous in its definition and explanations especially in the case of digital materials. But in the historic amendment of Indian Copy Right Act 2012, help the libraries to make the digitisation process easier without infringing the copyright Law. A new clause (n) has been introduced to enable the storage of a digital copy of a work if the library possesses a non-digital version of it. The amendment says:

*'(n) the storing of a work in any medium by electronic means by a non commercial public library, for preservation if the library already possesses a non-digital copy of the work;'* (<http://copyright.gov.in>)

While introducing technological protection measures, the amended law ensures that fair use survives in the digital era by providing special fair use provisions. The amendments have made many author-friendly amendments, special provisions for disabled, amendments facilitating access to works and other amendments to streamline copyright administration. The amendment also helps the library getting some special provisions for access to the disabled persons. The new clause (zb) added to section 52(1) providing for fair use of the work for the benefit of the disabled, facilitates adaptation, reproduction, issue of copies or communication to the public of any work in any accessible format, for persons with disability to access works including sharing with any person with disability for private or personal use, educational purposes or research. These rights are available to any person or organization working for the benefit of the persons with disabilities (<http://copyright.gov.in>).

On the basis of the above clauses, most of the intellectual property experts supported the digitisation of any documents in the library for preservation and non-commercial use. Although the landmark amendment has set a freedom for digitisation of library materials, in most of the library expert's opinion, a balance is to be drawn such that the rights of the publishing houses as well as the interests of the students are protected simultaneously without amounting to copyright infringement. On the light of the above mentioned amendments, the library has been granted more freedom in selecting the materials of the second phase of the digitisation project which was started on 5th October 2016.

### 11. Digitisation of Old and Rare Documents Phase-II

With the success of the Phase 1 of the digitisation, The Government of Kerala allotted an amount of 50 lakhs specifically for the project of digitisation of old and rare document phase 2 and the work was started on 5th October 2016, by M/S Informatics Ltd, Bangalore. The present status of the project is given in Table 2.

Total pages selected for digitisation	Total number of pages already scanned uploaded	Total number of books uploaded	Total number of periodicals scanned
10 Lakhs	479154 (as on January 2018)	1541	91

Table 2. Status of the Digitisation Project Phase II

### 12. Data Analysis and Discussions

Analysis of data collected through the various tools is the ultimate step in research process. It is the correlation between the raw data and significant results that leads to a conclusion. A detailed analysis of the data and its interpretation is presented below in the form of tables and graphs.

#### 12.1 Type of Documents Digitized

Table 3 shows the number of documents selected for the digitization phase-1

S.No	Type of Documents	Number	Percentage
1	Books	1005	83.26
2	Periodicals	202	16.74
	Total	1207	100

Table 3. Documents selected for the project

Table 3 clearly shows that the digitized collection includes 1207 documents comprising of 1005 (83.26%) books and 202 (16.74%) periodicals. Most of the documents are related to Kerala including its history, anthropology, arts, science, culture and literature.

### 12.2 Language-wise Distribution of Documents

Language-wise distribution of documents selected for digitization project is given in Table 4.

Language	Books		Periodicals		Total	
	Number	Percentage	Number	Percentage	Number	Percentage
English	581	57.81	38	18.81	619	51.28
Malayalam	260	25.87	156	77.23	416	34.47
Malayalam with Other languages	164	16.32	8	3.96	172	14.25
<b>Total</b>	<b>1005</b>	100.00	<b>202</b>	100.00	1207	100.00

Table 4. Language-wise Analysis of Documents

It is clear from the Table.4 that, 51.28% of documents are written in English, 34.47% in Malayalam and 14.25% in combination of Malayalam with other languages. Document-wise analysis shows that books chosen for digitization mostly written in English (57.81%) whereas periodicals selected for digitization mostly written in Malayalam language (77.23%). Graphical representation of the language-wise analysis of the digitized documents is given in Figure. 1

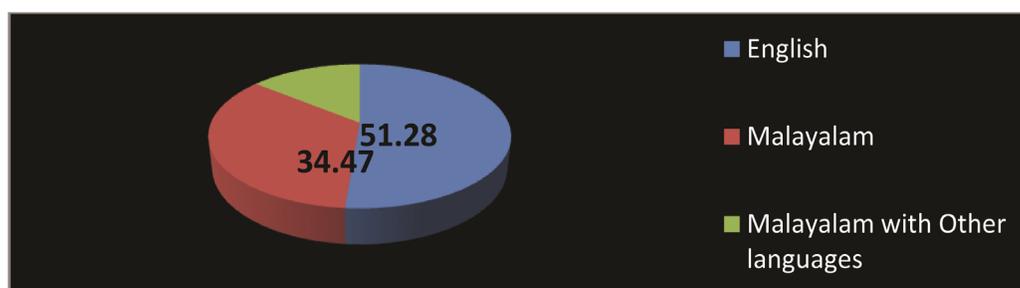


Figure 1. Language-wise analysis of documents

### 12.3 Subject-wise Distribution of Digitized Documents

Subject-wise distribution of digitized documents in the Kerala University Library is given in Table 5.

Subject	Books		Periodicals		Total	
	Number	Percentage	Number	Percentage	Number	Percentage
Multi-disciplinary	4	0.40	137	67.82	141	11.68
Archeology	8	0.80	3	1.49	11	0.91
Art & Architecture	13	1.29	3	1.49	16	1.33
Basic Science	47	4.68	1	0.50	48	3.98
Applied Science	24	2.39	1	0.50	25	2.07
Economics	85	8.46	7	3.47	92	7.62

Education	38	3.78	1	0.50	39	3.23
Geography	32	3.18	0	0.00	32	2.65
History	165	16.42	17	8.42	182	15.08
Language and literature	126	12.54	15	7.43	141	11.68
Law	75	7.46	2	0.99	77	6.38
Philosophy	11	1.09	1	0.50	12	0.99
Political science	280	27.86	3	1.49	283	23.45
religion	17	1.69	9	4.46	26	2.15
Sociology	80	7.96	2	0.99	82	6.79
<b>Total</b>	1005	100.00	202	100.00	1207	100.00

Table 5. Subject-wise Distribution of digitized Documents

Subject-wise analysis shows that 23.45% of the digitized documents belong to Political Science, 15.08% History documents, 11.68% in multi-disciplinary and 11.68% from language and literature. Document-wise analysis shows that more books are from Political Science filed (27.86%), followed by History (16.42%) and language and Literature (12.54%) whereas more periodicals are in the multi-disciplinary areas (67.82%) followed by History (8.42%) and Language and Literature (7.43%).

#### 12.4 Year-wise Distribution of Digitized Documents

Year-wise distribution of digitized documents in the Kerala University Library is given in Table.6.

Year	Books		Periodicals		Total	
	Number	Percentage	Number	Percentage	Number	Percentage
1800-1850	7	0.70	0	0.00	7	0.58
1851-1900	103	10.25	18	8.91	121	10.02
1901-1950	840	83.58	95	47.03	935	77.46
1951-1980	55	5.47	89	44.06	144	11.93
	1005	100.00	202	100.00	1207	100.00

Table 6. Year-wise Distribution of digitized Documents

It is clear from the Table that most of the documents (77.46%) were published during the year 1901-1950. It is also shows that 11.93% of documents were published during 1951-1980 and 10.02% during the periods of 1851-1900. Document-wise analysis clearly shows that most of the books (83.58%) and periodicals (47.03%) were published during the period of 1901-1950.

#### 13. Findings of the Study

Major findings of the study are:

1. Digitized collection includes 1207 documents comprising of 1005 (83.26%) books and 202(16.74%) periodicals.
2. Most of the documents are related to Kerala including its history, anthropology, art, science, culture and literature.

3. 51.28% of documents are written in English, 34.47% in Malayalam and 14.25% in combination of Malayalam with other languages.
4. Document-wise analysis shows that books chosen for digitization mostly written in English (57.81%) whereas periodicals selected for digitization mostly written in Malayalam language (77.23%).
5. Subject-wise analysis shows that 23.45% of the digitized documents belong to Political Science, 15.08% History documents, 11.68% in multi-disciplinary and 11.68% from language and literature.
6. Document-wise analysis shows that more books are from Political Science filed (27.86%), followed by History (16.42%) and language and Literature (12.54%) whereas more periodicals are in the multi-disciplinary areas (67.82%) followed by History (8.42%) and Language and Literature (7.43%).
7. Most of the documents (77.46%) were published during the year 1901-1950.
8. It is also shows that 11.93% of documents were published during 1951-1980 and 10.02% during the periods of 1851-1900.
9. Document-wise analysis clearly shows that most of the books (83.58%) and periodicals (47.03%) were published during the period of 1901-1950.

#### 14. Conclusion

Kerala University Library possesses a large number of old, rare and unique documents which are of the great value to the users especially the researchers. Those documents which do not come within the purview of copyright law were digitised and uploaded in the LIDAS, software for digitisation of the Kerala University Library. The Library faced lot of obstacles and challenges during the implementation of the project. Lack of proper digitization policy at the University level is one of the most important problems for starting the digitisation of old and rare documents. In order to solve this problem, UGC have to come forward to formulate a common policy for digitisation and coordinating the digitisation activities taking place at various Universities in India. As the first University level digitisation project of old and rare documents in Kerala, the Digital Archive of the Kerala University Library was a big success both in terms of their quality and quantity of the information. The digital archive is now widely used by the scholar not only from Kerala but from various other states and even from abroad.

#### References

- [1] Bultmann, B., Hardy, R., Muir, A., Wictor, C. (2006). Digitized content in the UK research library and archives sector. *Journal of Librarianship & Information Science*, 38 (2) 105- 122.
- [2] Indian Copy Right Act. <http://www.copyright.gov.in/Documents/Copyrightrules1957>. Retrieved on 12th February 2018.
- [3] Natarajan, M., Bhakta, U. (2008). *Developing a framework for digitisation: Issues and problems*. Paper presented at the National Conference on Recent Technological Trends in Management and Library System: Issues and Challenges, Asia- Pacific Inst of Management, New Delhi [http://knowgate.niscair.res.in/jspui/bitstream/123\\_456789/78/3/developing%20a%20frameork%20for%20digitisation](http://knowgate.niscair.res.in/jspui/bitstream/123_456789/78/3/developing%20a%20frameork%20for%20digitisation)
- [4] Mills, Alexandra. (2015). User Impact on Selection, Digitization, and the Development of Digital Special Collections, *New Review of Academic Librarianship*, 21. 60–169, 2015.
- [5] Rafiq, Muhammad. (2014). Towards a Digitization Framework: Pakistani Perspective *Journal of Information Management & Libraries*, Vol. 15.
- [6] Ringel, Sharon., Ribak, Rivka. (2015). Practicing Digitization at the National Library of Israel, *New Review of Information Networking*, 20. 236–240.
- [7] Tay Pek San, Peng, (2017). The impact of copyright law on the digitization of library collections in academic libraries in Malaysia. *Malaysian Journal of Library & Information Science*, 22 (1).
- [8] Sudhier, K. G. (2007). Digital archives of manuscripts in the ORI & MS Library, University of Kerala: A case study, *In: IASLIC All India Conference, JamiaMilliaIslamiya/sp.pub, No.47; 2007, p 213-218.*