

Noolaham Foundation Final Report



Project Title	Noolaham Digital Library Collection Development 2013
Project Number	NF/PG/2013/0001
Project Location	Sri Lanka
Sector	Collection Development, Digital Archive
Implementing Agency and contribution	Noolaham Foundation (NF)
Grant Agency and Contribution	Noolaham Foundation
Total Budget	N/A
Start date	1 January 2013
End date	31 December 2013
Responsible Stakeholder/s	Mr Rajathurai Ganesarajah, Mr. A. Jinnah Sherifdeen, Mr. N. Tharmarasa (Akalanagan) , J/Union College (Jaffna), J/Canagaratnam Madhya Maha Vidyalayam (Jaffna), and Local publishers of Sri Lankan Tamil speaking community

Summary

With the increased availability of economical digital storage media, high speed scanners and high-bandwidth networks, Digital Libraries have received a boost in the last few years. The dream of digitizing vast areas of knowledge pertaining to mankind, and making them available online has now become a realizable goal. The **Noolaham Digital Library Collection Development 2013** project focused on strengthening Noolaham's Digital Library collection by adding various rare documents to it.

Through this project Noolaham Foundation has digitized 676 documents and made them available on its Digital Library (www.noolaham.org). The documents that are made available online through the project date back to 1970. Most of these documents are out of print and are available at only a few libraries and personal collections.

Introduction and Background

Sri Lankan Tamil speaking communities hold knowledge as their primary and proudest resource. Learning is intertwined with our way of life. That is why upon hearing about the burning of Jaffna Library the communities were heartbroken. The loss, grief and sorrow that fell upon us will last for generations. We lost parts of our past and parts of our future.

Unfortunately, the burning of Jaffna Library was not the first tragedy. Nor was it the last. Historically, due to casteism, patriarchy, marginalization, colonization and other oppressive forces, knowledge and cultural ethos of many Sri Lankan Tamil speaking communities such as that of Plantation Tamils, Coast Veddhas and Tamil-speaking Muslims have not been fully recorded or shared. More than three decades long civil war, repeated displacements, migration, modernization and lack of awareness among the community have accelerated the loss of intellectual properties.

Research reveals that students, researchers and general peoples look online first when they need information because of the speed and convenience made available by online sources. They prefer remote access to electronic resources rather than having to go to a traditional physical library. Most of the peoples often turn to a digital library web site or electronic resources when they need information. Students and general people tend to start with popular Internet search engines like Google because these search engines are more convenient and easier to use than libraries. Most of the peoples believe the information they find on the Internet is good enough for knowledge building. Unfortunately, only about 6% of the surface web content indexed by popular search engines is appropriate for students' academic work. Scholars, researchers and teachers are concerned that the lack of quality resources on the surface web is having a negative impact on the quality of people's virtual learning.

As this project provides online access to a wide range of publications including monographs, magazines and journals, newspapers and other publications, all the Tamil speaking communities around the world shall benefit from this project. This project will also benefit individuals and institutions who want to digitally archive their publications. Thus this project will help the communities and institutions to preserve and share their knowledge bases and memory for the long term.

Implementation Methodology

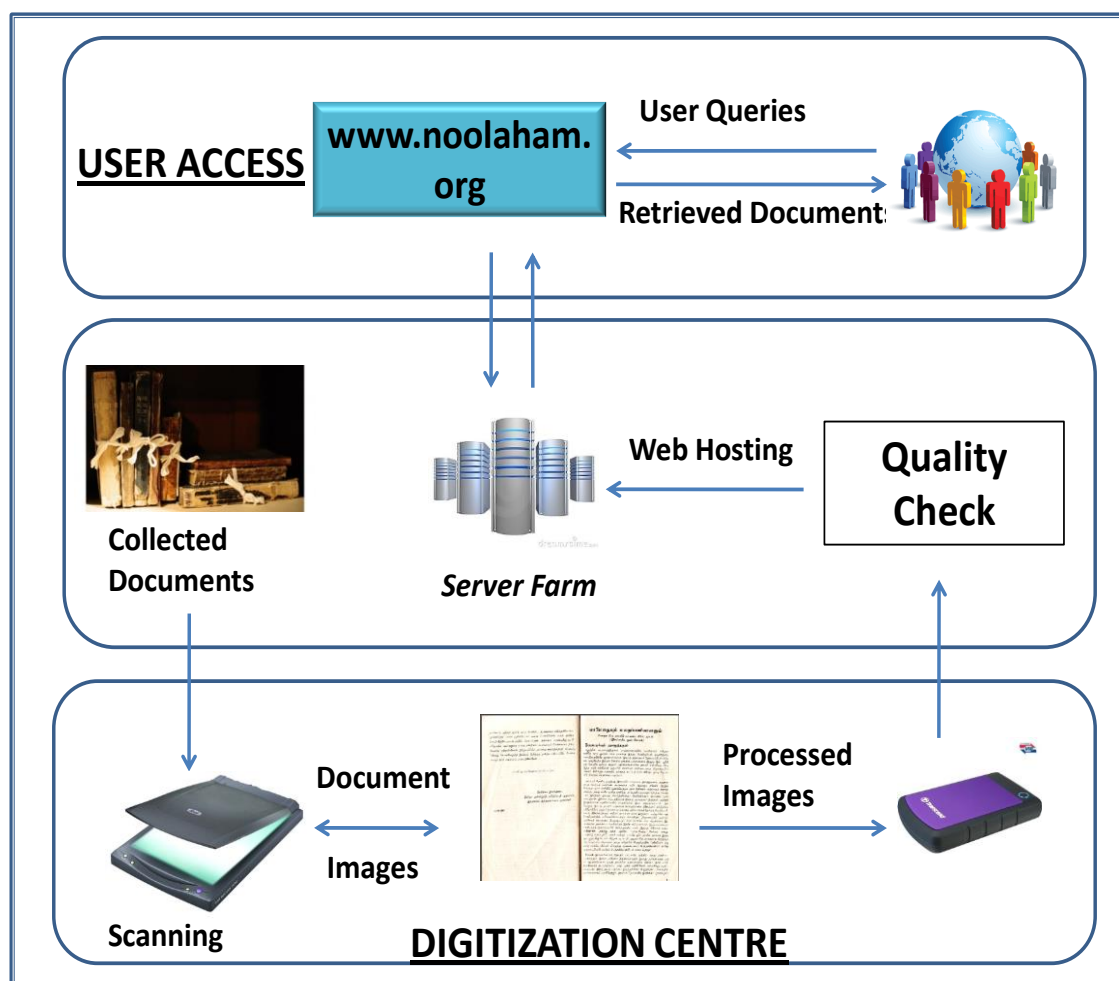
The entire process of digitizing the book consists of various stages such as procuring, scanning, creating Meta data, quality checking and web hosting. Noolaham Foundation has identified different phases of work that can be distributed over different locations. Owing to the magnitude of the project, the digitization process is distributed into different logical steps and over different staffs within a given digitization center.

The digitization process begins with the procurement of documents. A special trained human resource creates the metadata for the documents, which is checked for possible duplication against previously digitized documents. NF has built technology and solutions to avoid duplication within our center with minimal network resources.

The documents are collected from Mr. Rajathurai Ganesarajah (403 documents - Personal Collection Archive), Union College, Jaffna (9 books - Institutions Archive), Canagaratnam

Madya Maha Vidyalayam, Jaffna (7 books - Institutions Archive), Mr. Jinnah Sherifdeen (Personal Archive) - 9 books, Mr. N. Tharmarasa Akalangan, (36 books - Personal Archive), Contemporary Newspapers (99 issues) and Contemporary Magazines (113 issues) from various publishers from Sri Lanka's Tamil Speaking Communities.

Written permission was obtained from author or the publisher for copyrighted books to make them available online. Apart from printed books, a large number of rare and endangered unpublished documents which are a part of the heritage of Sri Lankan Tamils are also being digitized for preservation.



Overview of the Digital Library Process

To ensure that proper quality is maintained by the digitization center, the digitally converted book is checked for quality. If the contents contain errors, they are re-scanned and/or re-processed. A copy of the document is stored as a backup in case of any hard disk crash. These documents are hosted on the server as well.

Range of Digitized content created

Aspects	Diversity
Total Documents	676
Pages	
Publication Years	1969-2013
Medium of documents	Paper
Type of documents	Printed, Handwritten
Language	Tamil, English
Subjects	Art, Architecture, Autobiography, Astronomy, Commerce, Religious, Economics, Science, Engineering, Geography , Law, Health, History, Literature, Novels, Contemporary magazines and newspapers, etc
Sources of Books	Libraries, Educational Institutions, Personal collections, Personal Archive

Objective/s and Achievements or Results

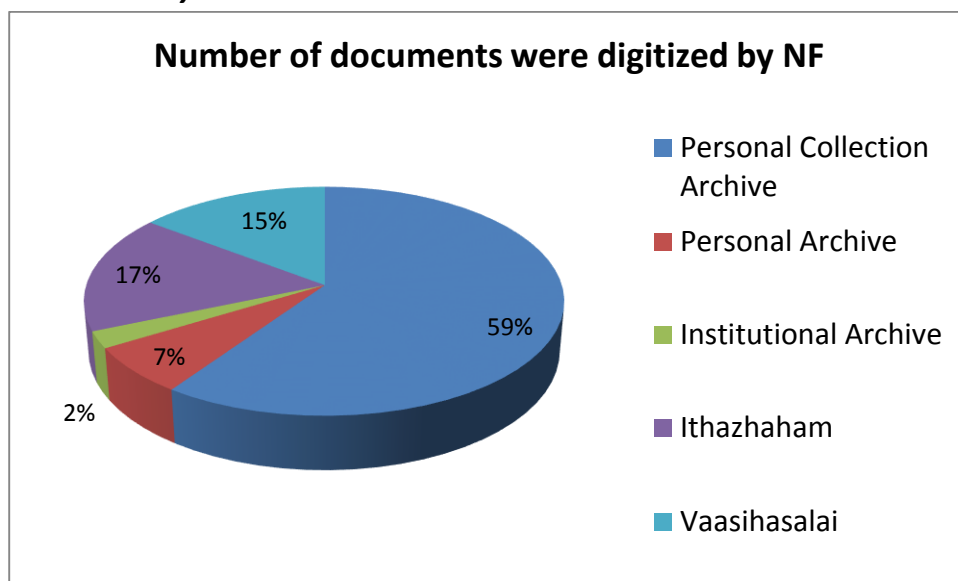
‘Noolaham Digital Library Collection Development 2013’ project offers strategic benefits for information sharing and preservation. It focused on strengthening Noolaham Foundation’s Digital Library (NDL) collections and digitized 676 publications related to Sri Lankan Tamil communities. Five of the nine collection areas of NDL newly acquired 676 publications, increasing the tally from 12,108 to 13,175 (The balance 391 documents came from Noolaham Foundation’s other projects).

Number of Documents by Category

Collections	Number of documents digitized
Personal Collection Archive	403 Documents
Personal Archive	45 Books
Institutional Archive	16 Books
Institutional Collection Archive	-
Ola-leaf Archive	-
Ithazhaham	113 Documents
Vaasihasalai	99 Documents
Malaiyaham Archive	-
Muslim Archive	-
Total	676 Documents

The overall objective of this project is to develop and strengthen the NDL collections and provide free and open access, thereby supporting educational, research and development endeavors related to Sri Lankan Tamil speaking communities. Also, specific objectives of the project include updating the institutional repositories of the organizations which have given permission to digitize their publications and further updating personal repositories of the individuals who have given permission to digitize their publications

Documents By collections



Constraints / Challenges

- Recollecting intellectual, rare and endangered documents is the major challenge to the development of digital library collections.
- This project showed that it was possible to secure permission to digitize works still in copyright provided publisher stipulations were met. Although some publishers granted permissions freely, others wanted to limit the use to the web or indicated that they would granted permission a few years from now.
- Another area where huge amount of money is required is hiring human resources, training the new staff and making required equipment available.
- The type of documents could vary in a variety of ways such as the size of the book, quality of paper, clarity of print, font face used, type of binding etc. Old documents which are delicate and have deteriorated over time need extra care apart from requiring special routines to process them.
- Noolaham Foundation implements this project in various centers all over Sri Lanka. The storage and transfer of the data generated by each center is a major task. Each center generates tens of GBs of data every day. Since network transfer of such large quantities of data is neither feasible nor economical, it has to be physically moved in Hard Drives or DVDs. Pipelining the data within the various phases of the digitization process is also a serious tactical issue.
- Due to the fact that different libraries have a copy of each of the many documents, duplicate digitization occurs, which results in wastage of effort and thus valuable resources. To avoid this, there has to be a proper synchronization system between the various digitization centers and head office.
- Maintaining large amount of digital data on the web for access from anywhere in the world at any time is a huge challenge. Accurate search and quick retrieval of the digital content against user queries is a major research problem.

- Metadata management is an important aspect of the digital library. A major portion of the sources of documents in the project have metadata only in non-digital formats and they have to be entered manually. For the entry of the metadata of a document, Noolaham Foundation largely relies on the relevant staff or the accuracy and credibility. Sometimes, staff might not be well aware of the hierarchy and ontology of book classification. It may affect the entire project and more commitment is much needed.

Suggestion and Recommendation

- There is a need for legal permission on intellectual property to digitize the documents. Noolaham Foundation should continuously try to get permission and consent from the document holders by explaining the importance of the documents that need protection and their usefulness for historical documentation and research.
- Noolaham Foundation has to increase its budget to facilitate digital library planning and digitization works. These kinds of work should be in tandem with other related organizations such as libraries and information centers. The collaborating activity will support virtual re-unification of digital library resources.
- Digital Libraries are also a means of preservation of content for the use of future generations. Hence maintaining good quality in the digitization process is of utmost importance. Also to ensure that all centers adhere to a common set of guidelines, enabling and maintaining a standard for the entire project is essential.
- Since no commercial OCR is available for Tamil language character recognition, the textual content is not available for these documents. This severely handicaps searching, which is restricted only to the title and keyword level. Special techniques need to be developed to search within the documents.
- Thus far, all the documents were only scanned in the binary mode. This is valid because most of the documents scanned do not have much colour content. Digitizing documents in colour would require more sophisticated scanners and better image processing and recognition algorithms and systems.
- The search engines of today work on the principle of keyword matching and perform searches in one language at a time. With a large corpus of multilingual data provided along with multilingual summarization and translation tools, a well-directed research effort would be needed to ensure concept- and content-based retrieval of knowledge from across multilingual data.
- A national level mechanism is essential to promote and coordinate open access and public domain digital library systems such as improving awareness of open access, regular training – tools, processes, standards, support setting up of working models, services

Some cover pages of digitized documents

One of "Akalankan" books and one of Mr. A. Jinnah Sherifdeen books

