

ADDIS ABABA UNIVERSITY DATAD PROJECT PROGRESS REPORT

1. Introduction

To improve management and access to African scholarly work, the Association of African Universities initiated the “Database of African Theses and Dissertations (DATAD) project”. The project was launched at the beginning of March 2002 at the Research and Graduate Programs Office of the Addis Ababa University. To this end, the Association of African Universities provided the necessary equipment (two Computers, one Printer and one Scanner) and the necessary software (OmniPage Pro, Procite, Word Perfect, Reference Web Poster, etc).

Since then, there has been progress in the conversion process of Thesis and Dissertations available in the Addis Ababa University. The report highlights the status of the work being done, in addition to the problems faced and their solutions, and proposals for improving implementation and, the way forward.

2. Actual number of theses and dissertations

Currently there are 2153 theses and dissertations available in our collection. Out of these, 51 theses are written in Amharic (the local language) local language and the rest (2102 theses and dissertations) in English.

3. The total number of records already indexed in the DATABASE

There were three steps in the conversion process: (1) scanning and recognition of the text by the OminiPage Pro OCR software, (2) classification of the softcopy data based on the indexing parameter used in the Procite database, and (3) editing entries in the database, including insertion of missing fields (such as keywords, gender, catalogue number, etc). Though there were challenges during the conversion process, the tables below shows the work done at each step.

Activity	No of theses	% (out of the total)
1. Scanned text/hard copy	2102*	97.6
2. Indexed records in the Procite database	2102	97.6
3. Finalized entries in the database	931	43.5

* The remaining 51 theses are Amharic texts that need translation before conversion

4. Contributions of the University for smooth implementation of the project

Besides its willingness to participate in the project, the University has taken the following measures to facilitate the work:

- Assigning both technical and professional staff for the tasks identified, such that the staff on the project had taken institution's time during data conversion process.
- Providing an office with the necessary facilities.
- Solving problems encountered in the day-to-day project implementation.

5. The challenges that had to be dealt with in implementing the project

- ✓ The existence of poor quality theses with regard to quality of paper and problems related to the readability (loss of connectivity, the existence of noise, etc) of scripts to the OCR system. The problem with this is that the OmniPage Pro scanner cannot recognize the characters correctly. The solution to this problem is to manually correct mis-recognized or unrecognized characters before classifying records in the database.
- ✓ The existence of theses which have no abstracts, keywords, gender, etc. Actions taken to solve the problems: (i) keywords are identified by analysing the abstract and content of the theses; (ii) gender is stated from the names of the researchers and, in case of difficulty, individual's file is referred.
- ✓ The difficulty of getting organized and readily available catalogue number (identification) for each thesis. The measure taken is to collect the identification number of the theses from the catalogue cards available in the library (Institute of Ethiopian studies)
- ✓ The problems of converting theses written in Amharic Language. Because both Procite and OmniPage Pro software couldn't understand Amharic characters, the solution would be to translate the content to English language.

6. Suggestions to enhance implementation of the project

As indicated in section three, the activities are categorized into three steps considering the interrelationship of the different tasks undertaken to achieve the aim of the project. Then, for each task individuals are assigned. Before starting the work, hands-on training is given to each individual on how to operate both the hardware and software. It is then that the actual conversion is started. We thought that this facilitated the performance of the work as shown in the above table.

In addition we suggest the following:

- ✓ There is a need to design and implement standardized theses report writing format that is inline with DATAD data requirements. Researchers have to be urged to

include keywords, gender, etc. in their theses presentations. It is also important to urge researchers to submit their theses both in hardcopy and soft copy formats.

- ✓ Site touring is also another important method to undertake on-the-job training, where more experienced individuals can help the less experienced ones to troubleshoot Procite, OmniPage Pro and Web Poster software, and the procedures to follow in facilitating the project.
- ✓ More of such meetings are crucial to share experiences, to evaluate the progress of work, and to discuss problems and find common solutions, especially for those issues needing a change of procedures.