

Final Report

Phase 2

Virtual Regional Dissertation & Thesis Archive

August 31, 2006

Submitted to:

Texas Center Research Fellows Grant Program 2005-2006

Submitted by:

Fen Lu, MLS, MS – Automated Services, Killam Library

Jeanette Hatcher, MLS, MAPAFF – Special Collections & Archives, Killam

Library

Table of Contents

I. Background Information.....	2
II. Research Statement of Purpose	2
III. Necessity of Research	2
IV. Research Methodology	3
Encoded Archival Description (EAD).....	4
Collection Development Parameters	4
Implementation Methodology	5
V. Findings.....	6
Digitization	6
Encoding	6
Transformation and Delivery.....	8
VI. Conclusion	9
Recommendations.....	9
Future Phases	10

Currently our project can be viewed at:

<http://library.tamtu.edu/depts/sparc/theses/index.htm>

Table of Figures

<i>Figure 1 -- Custom mapping code in Java & support text-to- EAD DTD mapping</i>	<i>7</i>
<i>Figure 2 -- Snapshot of XSL file</i>	<i>8</i>
<i>Figure 3 -- HTML display using XSL file</i>	<i>9</i>

I. Background Information

Grant Program: Texas Center Research Fellows Grant Program 2005-2006

Project Title: Phase 2 -- Virtual Regional Dissertation & Thesis Archive

Research Team: Fen Lu, MLS, MS – Automated Services, Killam Library

Jeanette Hatcher, MLS, MAPAFF – Special Collections & Archives, Killam Library

II. Research Statement of Purpose

The research team has developed an innovative historical preservation resource by creating a Regional Dissertation & Thesis Archive. As part of this endeavor we sought to provide electronic access, via the Internet, to vital information for each volume that is placed in the archive. This information includes the Title/Approval page, the table of contents, and the abstract for each dissertation & thesis. This initiative was documented in order to provide useful information to other institutions interested in undertaking a similar project.

III. Necessity of Research

Currently the Killam Library has an exceptional resource of information for the border region in that of the TAMIU/LSU Thesis Collection. Housed in the Library's Special Collections & Archives, this collection consists of 600+ volumes, the majority of which provide snapshots of demographic studies in our region. The majority of these theses deal with various topics of research whose scope of study is limited to the border region. These topics range from myriad issues in K-12 education in South Texas, to the potential impact of an International Trade Data System on the southbound commercial border. Faculty & student researchers, as well as private citizens and government entities, use these resources. The Special Collections & Archives, in the course of proctoring, arranging and providing access to these materials recognized the need for the following:

1. The need to provide improved physical and intellectual access to these materials.
2. The need to produce an expanded reference resource of this kind, which incorporates works about our region written by researchers at *other* institutions.

In accordance with these observations, the Special Collections & Archives and Automated Services departments of Killam Library project was a collaborative, multi phase, two-tiered project. This project sought to develop a new research resource and provide improved physical and intellectual access to these materials by providing Internet access to a Finding Aid created in an Encoded Archival Description (EAD) format.

IV. Research Methodology

The focus of this report is only the second phase of this project; however, it is important to give an overview of the entire endeavor. Future directions for this project such as the possibility of moving from open source software to licensed software largely depends upon the outcome of each successive phase. This project included the following proposed method of research:

Phase 1 consisted of:

Tier 1

- The creation of the TAMIU/LSU Thesis Collection
- Receipt of TAMIU/LSU Theses
- Provision of separate shelving for thesis collection
- The initial purchase and receipt of dissertations & theses whose subject matter deals with regional issues and/or examines regional groups authored by students seeking to fulfill academic requirements at institutions other than TAMIU.

Tier 2

- The creation of a Finding Aid for the TAMIU/LSU Thesis Collection
- The creation of a 2nd generation Finding Aid (EXCEL format) for the Thesis Collection, which includes title, author, submission date, call number and statement of purpose information.

Phase 2 consisted of:

Tier 1

- The creation of a Regional Dissertation & Thesis Archive
- Review and selection of TAMIU/LSU theses appropriate for inclusion in the new virtual collection
- Purchase of additional materials for the archive

Tier 2

- The creation of an Encoded Archival Description Finding Aid for this archive
- Creation of a template for an Encoded Archival Description (EAD) Finding Aid
- Scanning of the Title/Approval page, the table of contents, and the abstract of selected TAMIU/LSU theses
- Formatting scanned dissertation/thesis information to conform to the EAD Finding Aid
- Provision of access to the EAD Finding Aid via the internet

Encoded Archival Description (EAD)

It is important to note that by choosing to create an Encoded Archival Description (EAD) format finding aid, we are in line with widely accepted standards and best practices of the archival community. By using EAD, TAMIU Special Collections will be able, in future, to contribute content to larger archival initiatives such as those at the University of North Texas or The University of Texas at Austin and thus gain a wider audience for our collections.

Collection Development Parameters

Special Collections supports a broad range of regional research. Materials in Special Collections are collected because they hold long-term, historical research potential; possess unique physical characteristics, such as binding, printing or are heavily illustrated; are seminal, original works in a relevant area of study; or are inherently rare or scarce. Materials acquired can vary considerably according to the intrinsic qualities of the material itself, or the research needs of a particular program supported by the Special

Collections & Archives. The purpose of Special Collections is to preserve, conserve and provide access to these special materials.

Defined Project parameters:

- Subject Matter – Works dealing with border populations and/or issues including (but not limited to) art, culture, architecture, business, government, science, education and history
- Focus (but not limited to) -- The geographic region of South Texas, particularly the area of both sides of the Texas-Mexico border with a special emphasis on Laredo, Texas and Nuevo Laredo, Tamaulipas and the environs
- Language parameters – English and Spanish are the primary languages for this collection
- Format -- Dissertations, Theses, and Occasional Papers that have been *accepted* in partial fulfillment of requirements for completing a graduate course of study

Implementation Methodology

The project was implemented in three steps: digitization, encoding and delivery. It began with the digitization of 600 items selected from current TAMIU/LSU theses archives. Scanning technology was used to convert materials to a digital format. In order to observe copyright guidelines, only the cover page, abstract and table of content of these items were scanned. Parts of the current unpublished Excel format finding aid for the TAMI/LSU Thesis Collection were used for our project. To facilitate the implementation of EAD, the finding aid structure and content was standardized. The finding aid was encoded using Extensible Markable Language (XML) conforming to EAD definition version 2.0. During phase 2 of the project, open source software was used to perform the data conversion. To date, most Internet browsers are capable of reading XML-encoded documents but not all of them are. An XSLT file was developed to transform XML to HTML. Project staff mapped out the contents and navigation needs of the digital archive's Web page.

V. Findings

Digitization

The project began with digitization of 600 items selected from current TAMIU/LSU theses archives. Because we do not have any agreements with authors regarding the distribution of their theses and dissertations, we decided to scan only *cover pages*, *table of contents* and *abstracts* in order to avoid any potential copyright issues. In addition, all theses and dissertation were written in different formats. Even with the best scanning equipment, it would be very difficult to scan the table of contents and cover pages as text files and make them searchable. Therefore, we made the decision that the table of contents and cover pages were to be scanned as images and saved in PDF files. The abstracts were scanned as text and saved in separate files.

Student workers conducted the digitization process with two scanners. For three months approximately 29 hours each week was devoted to the scanning portion of this project. Student workers did the digitization following the procedure documentation, which defined the workflow, naming convention and all other relevant procedural details.

Encoding

Encoding is the most challenging part of implementation. The original inventory we created for our theses and dissertation collection is just a simple excel file in which each item in this collection is described with less than 20 features. Basically, it is just a table. However, EAD is a very complicated schema. It not only can present extensive and interrelated descriptive information found in finding aids, it also can preserve the hierarchical relationships that exist between different levels of description. The difficulty we faced is not just a simple changing of data from one format to another. The research team needed to change data from a format that contains no hierarchy to a format that has the potential to contain a very complicated hierarchy. Moreover, EAD also includes the descriptive information for finding aids. All kinds of pertinent information can be included. For example, the language being used to describe the collection or the type of character encoding being used in your finding aid can be included when using EAD.

Fortunately, the descriptive information within our finding aid is uniform. We can treat it as constant data and add it to our EAD XML file.

The following is the detailed description of implementation. First, in MS Excel, we exported our excel format inventory to text format. In this text file, each line represents a single row in the Excel file, i.e. a thesis/dissertation in our collection. Secondly, a Python program was developed to divide this text file into several files. Lastly, we downloaded Altova Mapforce 2006 trial version to generate a custom mapping code in Java and support the text-to- EAD DTD mapping (Figure 1). The Java codes generated by this application were used to convert all text files to XML files that conform to the EAD standard. We used the free version of Borland JBuilder Foundation to run the Java code.

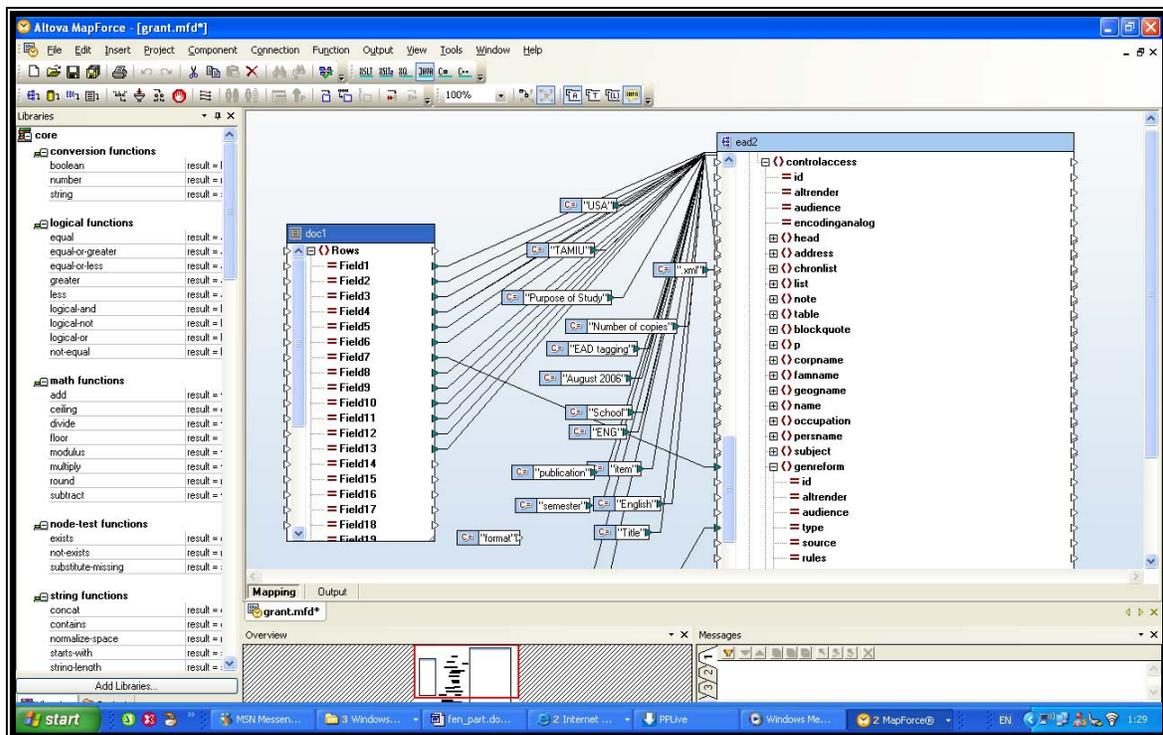


Figure 1

Custom mapping code in Java & support text-to- EAD DTD mapping

Curanderismo as an Example of an Emic Perspective in Contemporary Psychotherapy

Calderon, Yolanda Reynero
1995 Fall

Call#: Spec. Coll. Thesis Renero Calderon Yolanda 1995*

Subject: Counseling Psychology

- Committee ChairMadrigal, Julio
- Committee memberJohnson, Richard
- Committee memberRussell, Todd

Abstract:

Curanderismo is an ancient healing system derived from the Native American and Spanish folk traditions of Mexico. In the United States curanderismo is practiced mainly by traditional Mexican Americans who find solutions to their physical or mental-emotional problems in this method. The purpose of this study is to provide an introduction to the practice of curanderismo and its

[Table Of Content](#)

[Cover Page](#)

Browse

By Title
By Author Name
By Subject

Quick Links

Library Home
Special Collection Home
TAMU Home

Figure 3

HTML display using XSL file

VI. Conclusion

Recommendations

If other groups are contemplating starting a project similar to the one that this research team has undertaken, we have a few advisory comments to share with them. First, carefully consider the labor- intensive nature of this type of project and the skills that your workforce needs to possess in order to produce a viable product. Make sure that a knowledgeable and hardworking group is available for your project. Secondly, if you do not have the appropriate equipment at the outset of your project, remember to include extra time in your work timeline to acquire said equipment. Third, consider that once you create this database it can grow rather quickly. While free software can be appropriate for small or temporary databases and simple web interfaces, it may not be able to handle larger datasets or complicated search functions. Lastly, this was actually a multi-phase, multi-tiered project. Digitization projects of this kind take many months or years to be fully realized and require integration into the department's workflow.

Future Phases

There are several aspects that the research team is interested in exploring during future phases of this project. We are confident that with more research and more time, the process for the inputting of information into this database could be streamlined further; thus, allowing for a more efficient use of workers' time. The changes that we are going to explore will call for a shift in terms of training and our workers' necessary skill set. Streamlining input procedures will also allow for more time to engage in quality control activities.

Although at first it may seem superficial, the research team has spent some time discussing the visual representation of this project. As librarians, the research team understands the importance of having a visually engaging web interface. Naturally, utility of the tool is paramount in importance; however, it is extremely important to create an interface that is inviting and clearly mapped for the ease of users. In terms of aesthetics, the research team plans to continue evolving the look of the database's web interface. There are also plans to eventually provide users with a printer friendly page option.

There are also several possibilities in terms of increasing the utility of this tool. Namely by learning more about EAD descriptors and applying that knowledge to later versions of this database, the research team could greatly increase the usefulness of this product. Modifications of our current search interface and better use of EAD could greatly increase the search capabilities of this tool in later phases of this project.

It seems that much of the free software that was used during the earlier phases of this project will be useful for only a limited time. As the project grows, the research team will be under increasing pressure to purchase software and materials to continue development. Any improvements made in later phases will require increasing levels of technical expertise and be labor intensive. It appears that it will be necessary to seek future funding opportunities in order to fully develop later phases of this project.