

# HERE BE DRAGONS

## Navigating the Uncharted Waters of Legacy Thesis and Dissertation Digitization

### ISLAND OF PLENTIFUL PLANNING

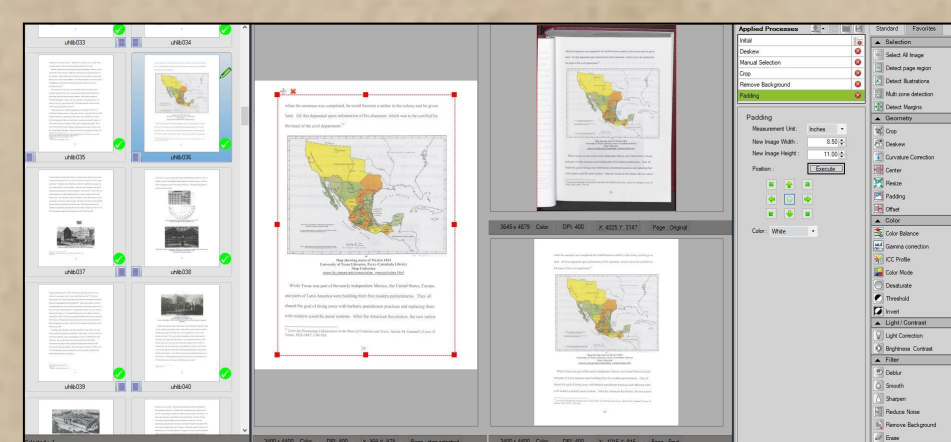
#### Phase 1: Preparation

##### ENVIRONMENTAL SCAN

- Evaluation of TDD project literature from 13 universities.
- Copyright considerations and the utilization of an opt-in, opt-out model were evaluated.
- Physical retention and preservation strategies were reviewed.

##### EQUIPMENT PURCHASES

- Two Canon DR-G2110 ADF Scanners
- Plustek Opticbook 3800L Flatbed
- Challenge Spartan 185A Paper Cutter
- LIMB Processing



Script testing on a thesis in LIMB Processing

##### INVENTORY & RETENTION

- One bound copy retained.
- Binding preserved on copies without duplicates
- Master spreadsheet used to track T/D.
- A team of student workers reviewed over 20K volumes for anomalies.

### OUTREACH OASIS

#### Phase 3: Preservation & Communication

##### PRESERVATION PLAN

- Reduce digital storage costs.
- 1 Master file (PDF) per volume is sent to preservation storage.

##### OUTREACH

- Promote content to Theses and Dissertations authors.
- Use e-newsletters and UH News emails to promote content to UH campus users.

TDD DSpace: <https://uh-ir.tdl.org/handle/10657/3932>

### WORKFLOW ATOLL

#### Phase 2: Procedures & Workflows

##### PULL VOLUMES & CATALOG DATABASE MAINTENANCE

1. Physical volumes are pulled by call number & moved to digitization shelves.
2. The volume records are withdrawn from the database.

OCLC #	Call	OCLC #	TITLE	Location	1984 Pulled Complete	1984 Digit Date	205 Reserve Status	761 Digit Complete
1007780	1983	Thesis 100 1983 P12	An experimental study of the effect of...	...	x			
1008817	1983	Thesis 100 1983 A10	An inquiry into the use of...	...	x			
1114000	1982	Thesis 100 1982 C107	A research study of the...	...	x			
1201006	1982	Thesis 100 1982 P10	...	...	x			
1304811	1982	Thesis 100 1982 P10	...	...	x			
1304812	1982	Thesis 100 1982 P10	...	...	x			
1304813	1982	Thesis 100 1982 P10	...	...	x			
1304814	1982	Thesis 100 1982 P10	...	...	x			
1304815	1982	Thesis 100 1982 P10	...	...	x			
1304816	1982	Thesis 100 1982 P10	...	...	x			
1304817	1982	Thesis 100 1982 P10	...	...	x			
1304818	1982	Thesis 100 1982 P10	...	...	x			
1304819	1982	Thesis 100 1982 P10	...	...	x			
1304820	1982	Thesis 100 1982 P10	...	...	x			
1304821	1982	Thesis 100 1982 P10	...	...	x			
1304822	1982	Thesis 100 1982 P10	...	...	x			
1304823	1982	Thesis 100 1982 P10	...	...	x			
1304824	1982	Thesis 100 1982 P10	...	...	x			
1304825	1982	Thesis 100 1982 P10	...	...	x			
1304826	1982	Thesis 100 1982 P10	...	...	x			
1304827	1982	Thesis 100 1982 P10	...	...	x			
1304828	1982	Thesis 100 1982 P10	...	...	x			
1304829	1982	Thesis 100 1982 P10	...	...	x			
1304830	1982	Thesis 100 1982 P10	...	...	x			
1304831	1982	Thesis 100 1982 P10	...	...	x			
1304832	1982	Thesis 100 1982 P10	...	...	x			
1304833	1982	Thesis 100 1982 P10	...	...	x			
1304834	1982	Thesis 100 1982 P10	...	...	x			
1304835	1982	Thesis 100 1982 P10	...	...	x			
1304836	1982	Thesis 100 1982 P10	...	...	x			
1304837	1982	Thesis 100 1982 P10	...	...	x			
1304838	1982	Thesis 100 1982 P10	...	...	x			
1304839	1982	Thesis 100 1982 P10	...	...	x			
1304840	1982	Thesis 100 1982 P10	...	...	x			
1304841	1982	Thesis 100 1982 P10	...	...	x			
1304842	1982	Thesis 100 1982 P10	...	...	x			
1304843	1982	Thesis 100 1982 P10	...	...	x			
1304844	1982	Thesis 100 1982 P10	...	...	x			
1304845	1982	Thesis 100 1982 P10	...	...	x			
1304846	1982	Thesis 100 1982 P10	...	...	x			
1304847	1982	Thesis 100 1982 P10	...	...	x			
1304848	1982	Thesis 100 1982 P10	...	...	x			
1304849	1982	Thesis 100 1982 P10	...	...	x			
1304850	1982	Thesis 100 1982 P10	...	...	x			
1304851	1982	Thesis 100 1982 P10	...	...	x			
1304852	1982	Thesis 100 1982 P10	...	...	x			
1304853	1982	Thesis 100 1982 P10	...	...	x			
1304854	1982	Thesis 100 1982 P10	...	...	x			
1304855	1982	Thesis 100 1982 P10	...	...	x			
1304856	1982	Thesis 100 1982 P10	...	...	x			
1304857	1982	Thesis 100 1982 P10	...	...	x			
1304858	1982	Thesis 100 1982 P10	...	...	x			
1304859	1982	Thesis 100 1982 P10	...	...	x			
1304860	1982	Thesis 100 1982 P10	...	...	x			
1304861	1982	Thesis 100 1982 P10	...	...	x			
1304862	1982	Thesis 100 1982 P10	...	...	x			
1304863	1982	Thesis 100 1982 P10	...	...	x			
1304864	1982	Thesis 100 1982 P10	...	...	x			
1304865	1982	Thesis 100 1982 P10	...	...	x			
1304866	1982	Thesis 100 1982 P10	...	...	x			
1304867	1982	Thesis 100 1982 P10	...	...	x			
1304868	1982	Thesis 100 1982 P10	...	...	x			
1304869	1982	Thesis 100 1982 P10	...	...	x			
1304870	1982	Thesis 100 1982 P10	...	...	x			
1304871	1982	Thesis 100 1982 P10	...	...	x			
1304872	1982	Thesis 100 1982 P10	...	...	x			
1304873	1982	Thesis 100 1982 P10	...	...	x			
1304874	1982	Thesis 100 1982 P10	...	...	x			
1304875	1982	Thesis 100 1982 P10	...	...	x			
1304876	1982	Thesis 100 1982 P10	...	...	x			
1304877	1982	Thesis 100 1982 P10	...	...	x			
1304878	1982	Thesis 100 1982 P10	...	...	x			
1304879	1982	Thesis 100 1982 P10	...	...	x			
1304880	1982	Thesis 100 1982 P10	...	...	x			
1304881	1982	Thesis 100 1982 P10	...	...	x			
1304882	1982	Thesis 100 1982 P10	...	...	x			
1304883	1982	Thesis 100 1982 P10	...	...	x			
1304884	1982	Thesis 100 1982 P10	...	...	x			
1304885	1982	Thesis 100 1982 P10	...	...	x			
1304886	1982	Thesis 100 1982 P10	...	...	x			
1304887	1982	Thesis 100 1982 P10	...	...	x			
1304888	1982	Thesis 100 1982 P10	...	...	x			
1304889	1982	Thesis 100 1982 P10	...	...	x			
1304890	1982	Thesis 100 1982 P10	...	...	x			
1304891	1982	Thesis 100 1982 P10	...	...	x			
1304892	1982	Thesis 100 1982 P10	...	...	x			
1304893	1982	Thesis 100 1982 P10	...	...	x			
1304894	1982	Thesis 100 1982 P10	...	...	x			
1304895	1982	Thesis 100 1982 P10	...	...	x			
1304896	1982	Thesis 100 1982 P10	...	...	x			
1304897	1982	Thesis 100 1982 P10	...	...	x			
1304898	1982	Thesis 100 1982 P10	...	...	x			
1304899	1982	Thesis 100 1982 P10	...	...	x			
1304900	1982	Thesis 100 1982 P10	...	...	x			

TDD Tracking System

##### DIGITIZATION WORKFLOW

1. **Batch Prep & Disbinding**
  - a. Volumes are pulled from shelving and transferred via book cart to a disbinding room.
  - b. Disbound volumes are filed according to OCLC numbers.
2. **Scanning**
  - a. Physical volumes are grouped in batches of 50. A text file is created with a list of the OCLC numbers in a batch.
  - b. A custom Ruby executable reads the text file and creates a scanning directory.
  - c. Volume is physically checked and scanned on feed or flatbed scanner.
3. **Image Processing**
  - a. LIMB processing is used to convert raw image data to files for PDF packaging.
4. **OCR**
  - a. ABBYY FineReader Server 14 is used to generate PDFs and OCR data.
  - b. A custom Ruby command line app is used for automated file management and quality control.

##### METADATA

1. **Metadata Standards & Edits**
  - a. Metadata crosswalk from MARC records to Qualified Dublin Core records (one time batch process).
  - b. Quality control - Evaluate each record, edit when needed.
  - c. Metadata enrichment - add abstracts, chairs and advisors.
2. **Copyright Verification**
  - a. Evaluate each volume's copyright status.
  - b. Assign out of copyright or in-copyright statement.

##### DOCUMENTATION

- a. Complete workflow is captured & updated in live document.
- b. Update Master spreadsheet as batches move through workflow.

19,000 DIGITIZED THESES  
& DISSERTATIONS

THESES & DISSERTATIONS DIGITIZATION (TDD) TASK FORCE, UNIVERSITY OF HOUSTON LIBRARIES, TEXAS CONFERENCE ON DIGITAL LIBRARIES 2021

Map images courtesy of University of North Texas Libraries, Portal to Texas History | Graphics from Pixabay.com

##### TDD CREW

Crystal Cooper Bethany Scott  
Taylor Davis Van-Atta Santi Thompson  
Jerrell Jones Anne Washington  
A. Laura Ramirez Andrew Weidner  
Annie Wu