HERE BE DRAGONS
Navigating the Uncharted Waters of Legacy Thesis and Dissertation Digitization

ISLAND OF PLENTIFUL PLANNING
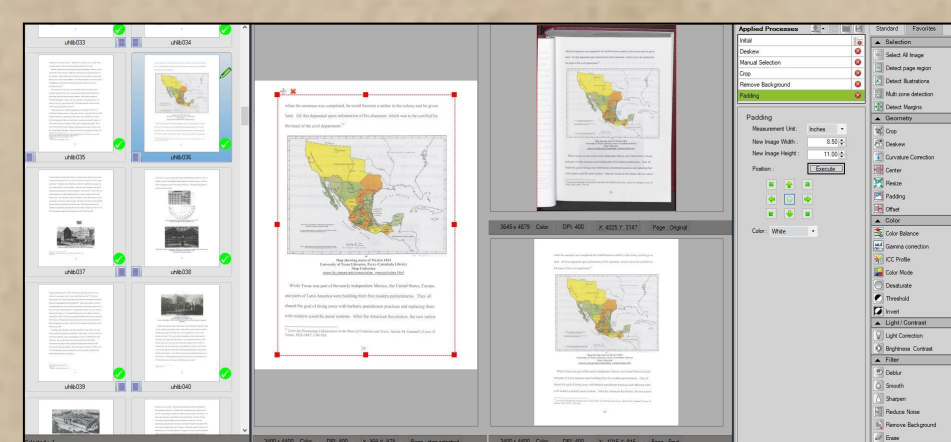
Phase 1: Preparation

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

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.

Script testing on a thesis in LIMB Processing

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 The volume records are withdrawn from the database.

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.

REPOSITORY INGEST
1. **File Ingest**
   a. Use command line to create CSV file.
   b. Use SAF Creator to create SAF ingest package, import to DSpace.
2. **Access Control**
   a. Use Selenium IDE to restrict access of **in-copyright** volumes to UH users.
   b. All metadata remains accessible.

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

19,000 DIGITIZED THESES & DISSERTATIONS

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