

The Board of Regents reports and minutes: a digitization case study

Jeremy Moore, Megan North, Todd Peters, Erin Mazzei
Digital & Web Services Department and University Archives, Albert B. Alkek Library, Texas State University

Necessary Background Information

Student imaging technicians at Alkek Library's Digital & Web Services Department at Texas State University digitized the University Archives' Board of Regents reports and minutes in Spring 2017.

Collection Scope

The Board of Regents was established in 1911 by the Texas Legislature and is responsible for the general control and management of the universities in the system. The Board of Regents reports and minutes is an ongoing series, and contains reports submitted by each school in the system, and minutes from the quarterly and special called meetings.

Collection Size and Materials

- 25 linear feet totaling 53,004 pages
- 51,229 pages in 225 bound volumes we disbound
- 728 pages in 18 bound volumes scanned bound
- 1,047 onion skin and loose pages in 21 folders

Onion Skin & Bound Volume Scanning

Initially Unacceptable Results with Plustek A300

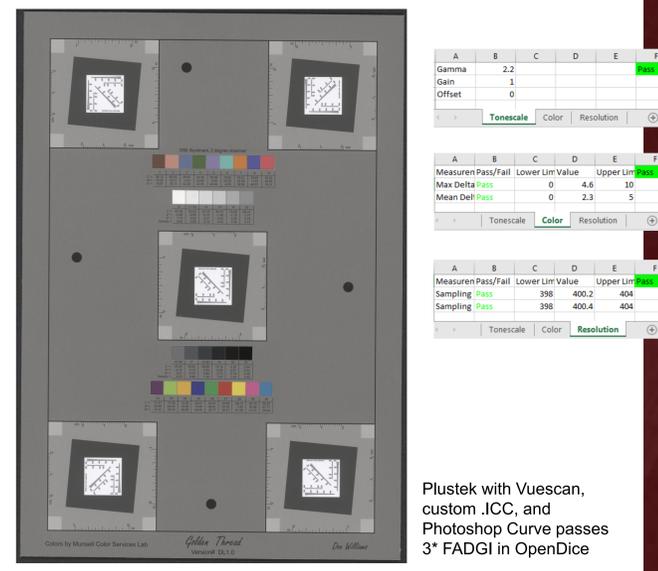
- OEM software ALWAYS auto-levels scanned area
- Color balance and tonal response varies per scan
- Not enough toggles and options to alter scans
- Scans at 400 ppi effectively resolve 265 ppi



Plustek A300 & VueScan (Hamrick.com)

Steps to FADGI 3* with VueScan

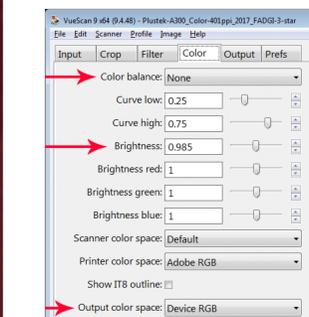
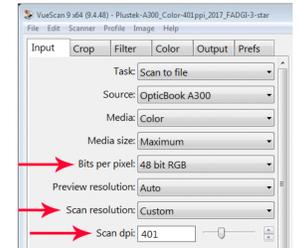
- No auto-levels = consistent output & custom .ICC use
- Custom .ICC created with BasICColor Input 5
- Custom Photoshop curve applied
- Scans at 401 ppi effectively resolve within 3* specs



VueScan & Photoshop Settings

VueScan Input Options

- 48 bit RGB
- Resolution: Custom
- Scan dpi: 401

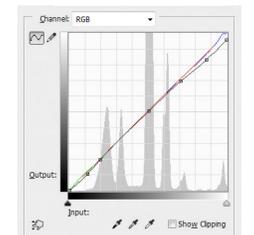


VueScan Color Options

- Color balance: None
- Brightness: 0.985
- Output color space: Device RGB

Photoshop Curve Adjustments

- Reduce highlights
- Darken shadows
- Increase blue in highlights
- Increase green in shadows



Why Digitize This Instead of That?

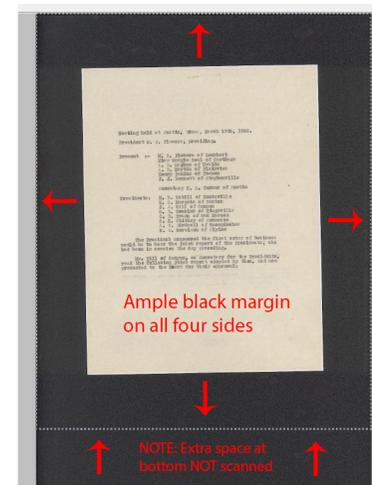
Many offices on campus hold original bound issues; making these available digitally and searchable should be a valuable service to the University administration.

- Use is more easily tracked with web analytics
- Self-service access frees up University Archives' staff
- Wide usage may result in new digitization partnerships and projects across campus
- Once digitized can be moved to off-site repository
- Student imaging technicians already trained to guillotine bound volumes and scan them in a duplex sheet-feed scanner
- *Imaging techs needed and wanted more to do!*

Easy Scanning & Automation AKA How We Prevent Carpal Tunnel Syndrome

Easier Scanning = Faster Scanning = More Attention to Materials' Safety

- Student imaging technician loads custom Vuescan preset: *Plustek-A300_Color-401ppi_2017_FADGI-3-star.ini*
- Pre-scans for scan area and sets a crop box in Vuescan software with wide black margins
- Scan and output documents to Network Share as maximum quality JPEG to speed upload
- Students focus on safely handling materials instead of clicking boxes while scanning



Automated processing

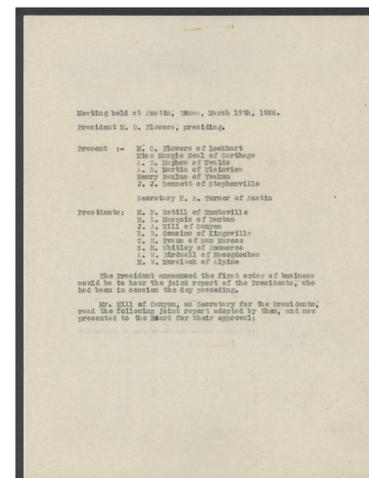
- Full Time Staff downloads JPEGs
- Custom .ICC, AdobeRGB, and Photoshop curve are applied with custom action
- File is saved as uncompressed TIFF

Autocrop files with Python 2.7 and openCV

- Full Time Staff runs custom Python script autoCrop_Plustek300.py from command-line
- Python script deskews document and crops file with pixel padding set at runtime
- Image at right cropped with a value of 50 pixels; negative values can also be used

Image File Properties and Time

- Input: 66mb TIFF; Output: 47MB TIFF
- Average crop time: 0.63 sec / image



Student-Led Duplex Scanning

Highly Trained Techs (combined experience: 6.25 years) Project management and workflow for scanning 51,229 pages from 225 guillotined volumes was left to student imaging technicians.

Process

1. Students guillotined then processed volumes into single pages
 - Removed staples and checked each page for glue
2. Scanned pages with Fujitsu fi-6670 Duplex Scanner
3. Processed pages with ImageMagick through Cygwin (Unix-like command-line for Windows) and Photoshop
4. Performed quality control over scanned volumes and created OCR'd PDFs for access

Faster, Cheaper, Better

Even with rescanning the onion skin pages using the new VueScan/Photoshop/Python Script process, we saved over 18 hours of work.

The power of code: write it once and run it over and over.

AutoCropping 10,000 scans will save 250 hours of time making for happier, more productive student technicians.

Acknowledgements

We want to thank and recognize our dedicated and highly talented Student Imaging Technicians:

Ashton Woodward Luke Sebree
Oscar Martinez Vicente Rangel

Time to Autocrop 1,047 files: 10 min, 57 sec
Average time to open, manually deskew and crop, then save : 1 min 30 sec/image
Time saved and boring work skipped with script: 1559.55 min or 25.99 hrs or 1.08 days