Extension Microfilm Digitization Project: Putting History Into Our Hands

Inga Haugen
Meagan Russell
Julia Westblade
01. Inga

Virginia Agricultural Publications Inventory
Virginia Ag Pubs Inventory (VAPI); materials from VSU, VT, and VCE

VSU - Virginia State University
VT - Virginia Tech
VCE – Virginia Cooperative Extension
Ag Pubs – any materials, artifacts, or items created in the course of our work at these institutions meant to be seen, used, and available to the public
02. Digital Imaging Lab

Virginia Tech’s Newman Library
The Digital Imaging Lab provides high quality digitization services in a wide variety of formats from across the library collections and other cultural heritage institutions. Highly trained staff combine state-of-the-art digitization equipment with best practices to provide support for collection-level digitization projects, systematic digitization, as well as on-demand requests by VT researchers, faculty, staff, and other library patrons.
Reel Digitization

• Create digitization guide
• Scan the microfilm
• Process/identify pages
• Straighten and crop pages
• Flip and/or color correct pages
• Identify and sort documents
• Rename files
• Create JPG derivatives
• Create searchable PDFs
• Pass off to metadata team
Create Digitization Guide

The Digital Imaging Lab creates a digitization guide for every project we do. We record basic metadata such as who did each step of the work and when. We also record any notes that we would like to pass along to the next steps of the project.
The lab digitized 141 reels of microfilmed from the Mekel scanner using Quantum Scanning software, then we pushed the images to the server and used the Quantum Processing software to identify individual pages.
Early reels were scanned one page at a time but part way through they changed to two. The pages were not always straight, and frequently every other page was upside down. We had to check every page’s orientation in the QC process.
Identify and sort documents

Meagan went through each reel and identified groups of images for each document. Some were broken into two parts. Students grouped the identified images into individual folders – one for each document.
**Identify and sort documents**

Things got complicated somewhere around 2/3 of the way through the process when the pages were not grouped at all and weren't scanned in any sort of order. We could either ...

- Value speed and break images into groups of 50 in the order they were scanned and add a note to researchers so that they could figure it out on their own.
- Spend the time to individually sort pages and provide as many complete documents to researchers as possible.

Where possible we chose option #2.
Rename files and create derivatives

Once the images were sorted into folders, we used a renaming tool and Python scripts to batch rename files according to our file naming convention.

We used Photoshop's Image Processor to then convert those folders of tifs into folders of jpgs.
Create searchable PDFs

Student workers in the lab used Adobe Acrobat to combine the jpgs into PDFs and apply OCR.

These are stored on the server in their own folder.

Current count is 10,989 PDFs.
Next Steps

...we scanned them. Now what?
Once it leaves the lab

Inga's team is working on metadata for each document.

They are collecting basic information including
• Document title
• Author/ person who compiled the report
• The county covered
• Date
Virginia Cooperative Extension (VCE)

Permanent URI for this community https://hdl.handle.net/10919/5523

Virginia Cooperative Extension is an educational outreach program of Virginia's land-grant universities: Virginia Tech and Virginia State University, and a part of the National Institute for Food and Agriculture, an agency of the United States Department of Agriculture.
Challenges

(Once upon a time we thought we'd only have 2000 PDFs)
Hardware, software, time constraints

• The scanner required some repairs and visits from technicians and workarounds.
• On some reels, every other page was blank - delete them for more coherent documents and smaller file sizes? Or leave them to save time?
• Some of the documents are in order but backwards.
• Students working simultaneously can cause confusion so we would have to be clear about which aspect or reel each student was doing. Meticulous documentation is helpful.
• The project is 5-6 times larger than we'd anticipated.
• Leadership of the lab changed part way through the project.
The quality of the film and original scans

The film was often scratched and lighting was inconsistent.
The quality of the film and original scans
OCR doesn’t always pick everything up

We just used the basic OCR feature in Adobe and for some pages, it was unable to detect some or any text.
Ransomware Attack

Totally normal challenge.

Early in the project, the library was part of the Kaseya ransomware attack and everything we had finished to that point was corrupted.

We were able to get all the images back but we lost some of our technical metadata.
Questions?  
Julia – juliags@vt.edu  
Inga – ihaugen@vt.eu  
Meagan – rmeagan9@vt.edu