Digitization of Texas Agricultural Experiment Station Documents for Inclusion in the Institutional Repository and World Wide Access

By
Dr. Rob McGeachin
Texas A&M University Libraries
October 9, 2006

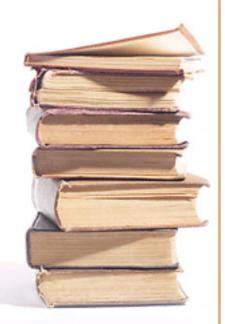
Overview

- Project Goals
- Project Design
 - Equipment & software
 - File naming
 - Image and PDF files
 - Metadata
- Institutional Repository



Project Goals

- Develop Ag Content for National Digital Library for Agriculture
- Archival Quality Digitization of the Bulletin of the Texas Agricultural Experiment Station
- Make Texas Publications Accessible to the World
- Make Metadata Records
 Harvestable to Other Indexing and Search Sites



Funding

- Specific Cooperative
 Agreement with the National Agricultural Library (USDA ARS)
- Three years -- \$11,990 for student wages and supplies
- Cost sharing of Principal Investigator's time – 20%



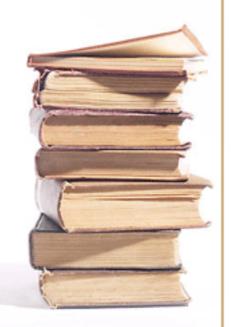
Project Design – Equipment and Software

- Computer Workstation
- 100 GB External Hard Drive
- OpticBook 3600 Book Edge Scanner and Software
- Adobe Acrobat Ver. 7.0
- D-Space Software
- Lots of IR File Storage about 600 GB



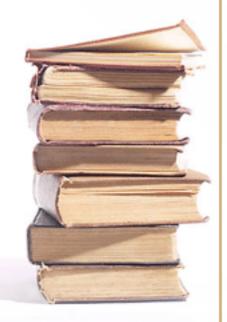
Project Design - File Naming

- Name folders by Publication and Publication Year
- Ex: TAES Bulletin
 - -1888
 - -1889
- Filename: Bulletin Number, Page Number
- Ex: b0001 0001.tif b0001 0004.tif;
- Filename: Bulletin Number
- ex: Bull0001.pdf Bull0535.pdf



Project Design – TIFF Image Files

- One image file per page, including blanks
- Scan at 400 dpi for text pages
- Scan at 600 dpi for illustration pages
- Scan as grayscale unless color illustration
- Text pages = 6 MB, illustrations = 14 MB



Project Design – PDF Files

- Add each page image file to document
- Perform OCR on all pages to generate text
- Rotate illustration pages 90 degrees
- Save as fully searchable PDF file
- Full Bulletin text searchable in TxSpace



Project Design -- Metadata

- NAL modified Dublin Core
- Include author, title, keywords, date, series
- Include NAL Thesaurus Subject Terms
- Initial Records in Excel Spreadsheet
- Entry part of D-Space workflow
- Metadata OAI harvestable



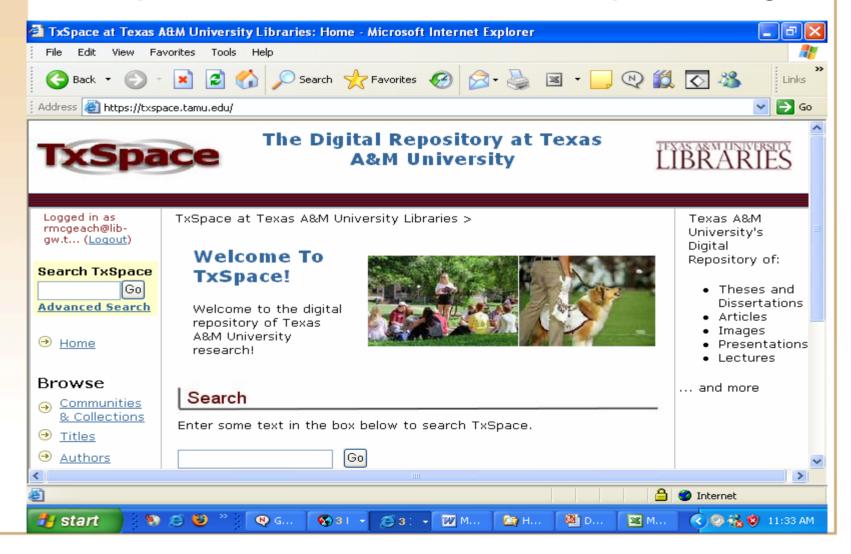
Institutional Repository

- TxSpace
- http://txspace.tamu.edu/
- Organized by Academic and Agency Units
- Upload Archival TIFF files, PDF Files, and Enter Metadata
- Example:

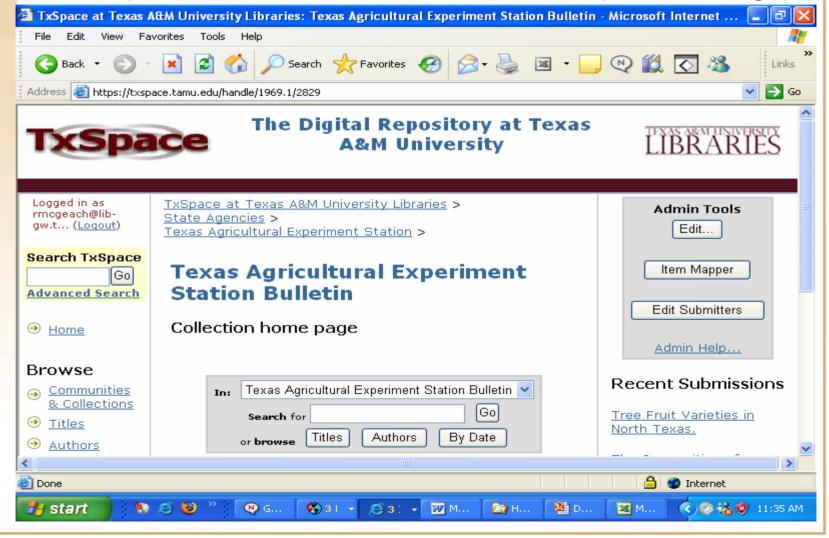
https://txspace.tamu.edu/bitstr eam/1969.1/2850/9/Bull0012.pd f

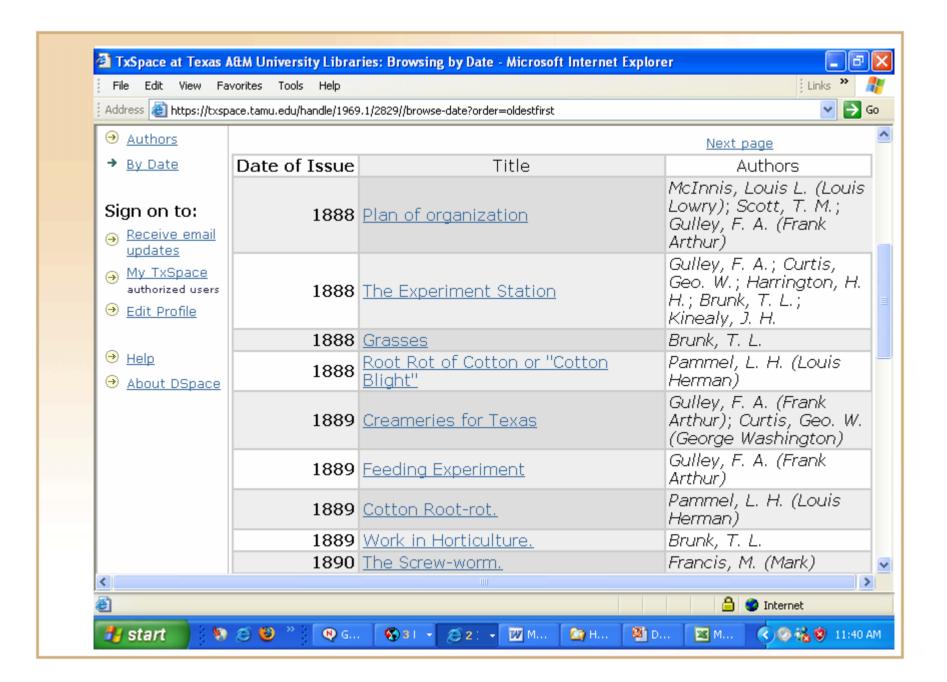


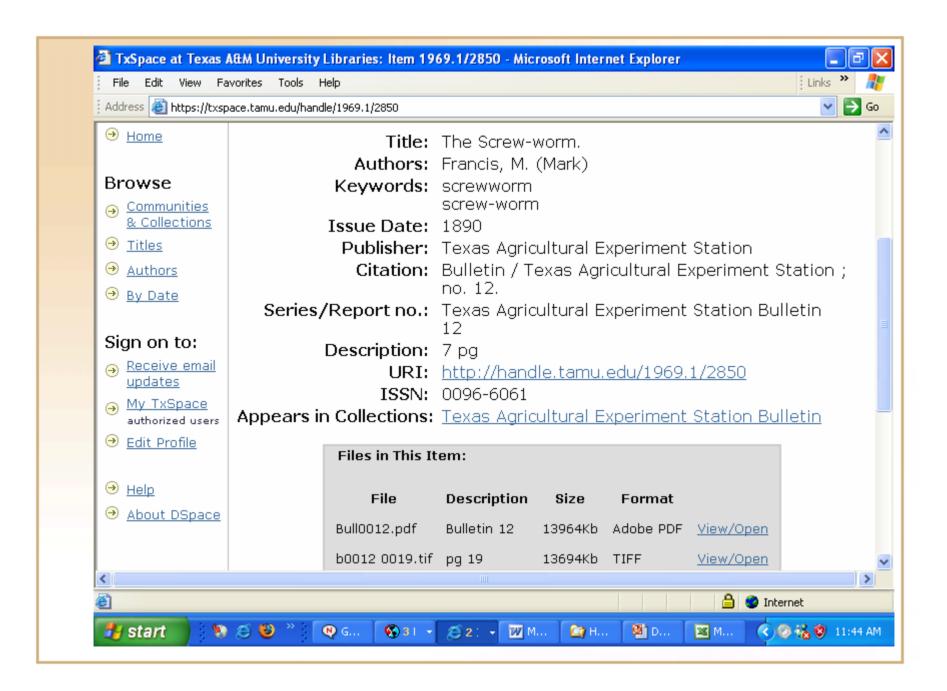
TxSpace Institutional Repository

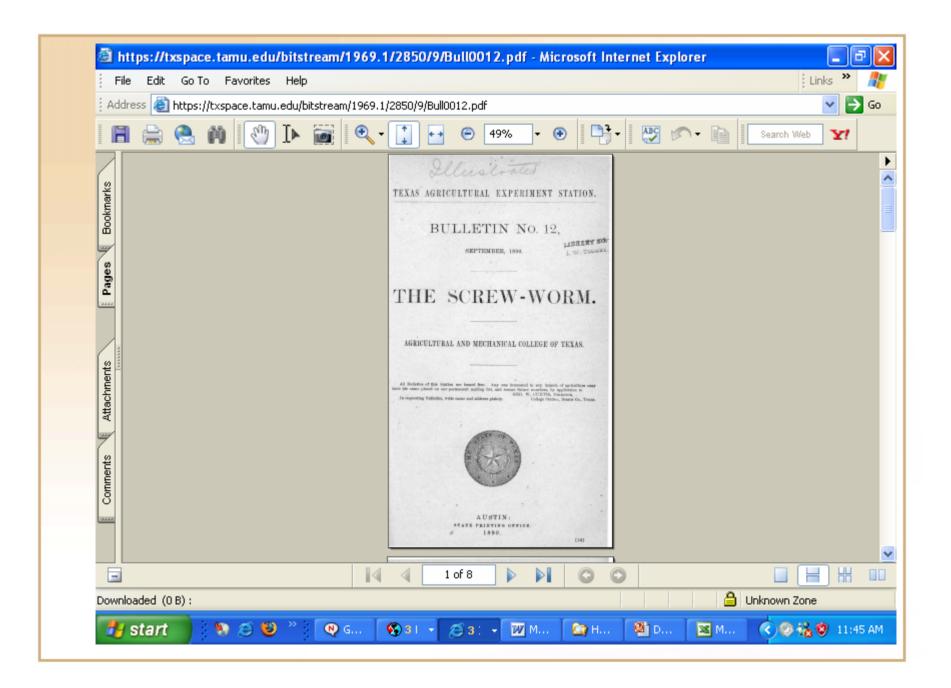


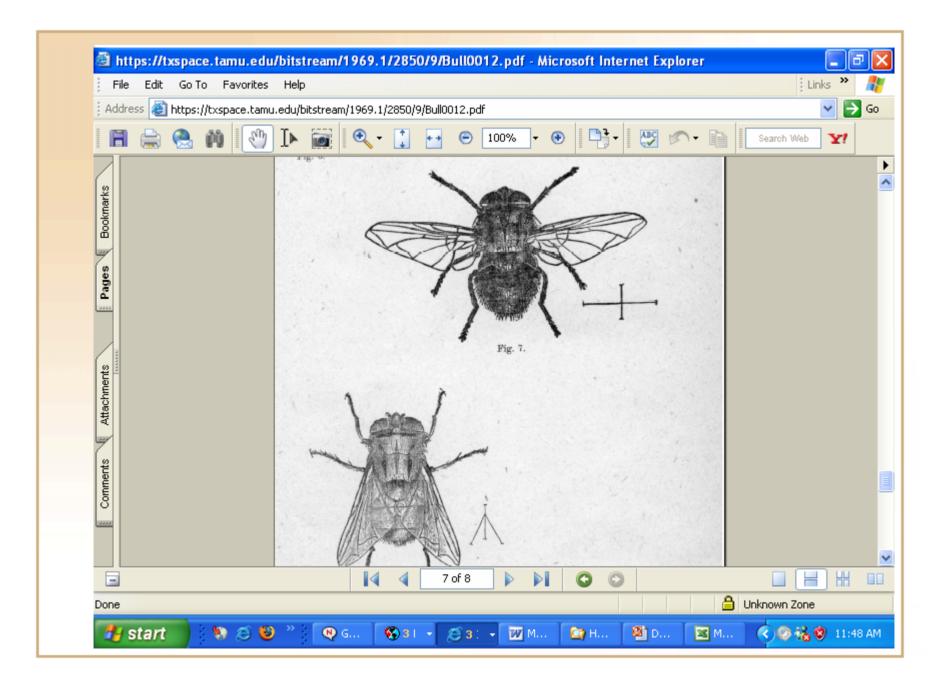
TxSpace Institutional Repository

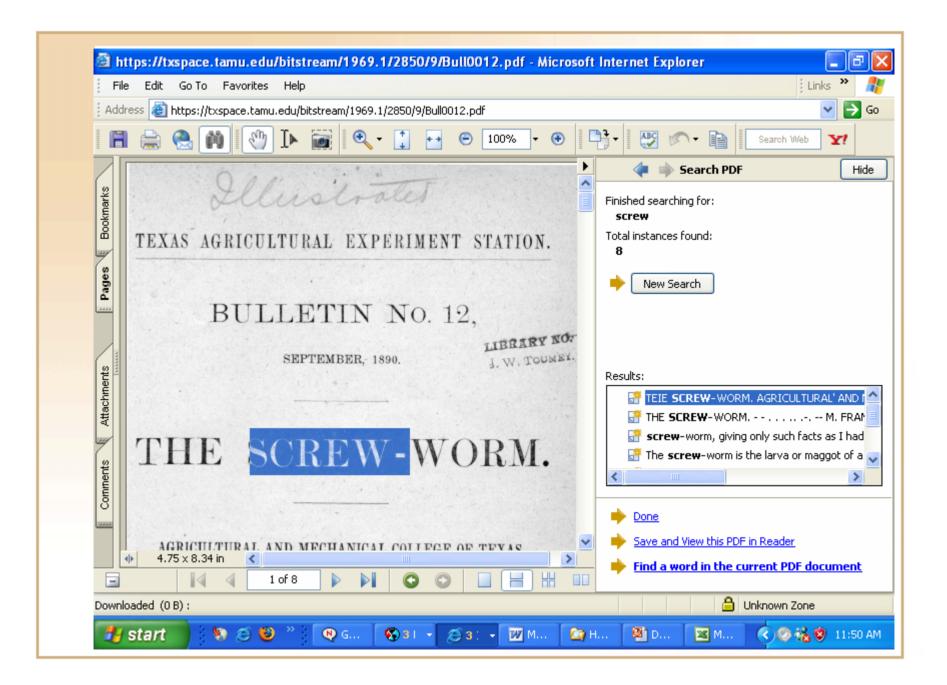








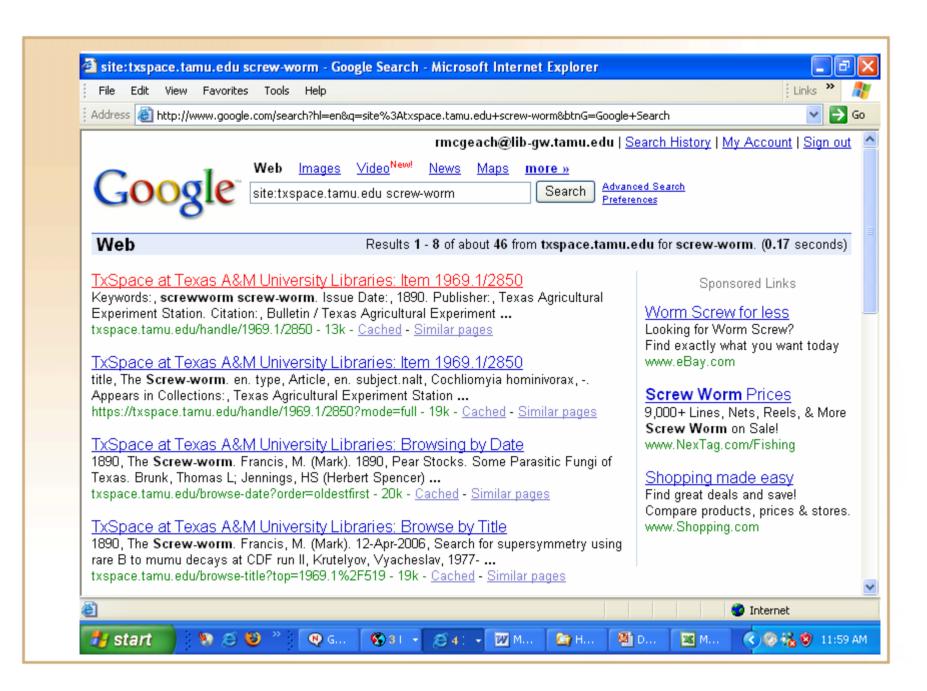




Google Search

 Site:txspace.tamu.edu screw-worm





This is a work in progress!

Any Questions?

