

Project Objectives

With funding from the National Historical Publications and Records Commission, the Railroad Commission of Texas sought to digitize approximately 171 linear feet or 4,623 folders that contain records of national historical significance from regulatory hearings that occurred between 1932 and 1972. Railroad Commission hearings concern various oil and gas exploration and development issues including field rules, secondary recovery projects, maximum efficient rates of production, determination of responsibility for the proper plugging of abandoned wells, applications to inject water into reservoirs for enhanced oil and gas production, and prevention and control of oil and gas pollution.

With the initial proposal, the Commission established several objectives for the second year of the grant period. The Commission anticipated that by year two all imaging and quality assurance work would be complete, the hearings files would be available online, and that much of the activity for the period would focus on preparing the finding aid, promoting the collection, and monitoring usage. Unfortunately, each phase of the project took much longer than originally anticipated, delaying the schedule by approximately 6 months. In December 2010, the Commission requested and received an extension to continue the project through June 2011.

The project has been much more complex than initially thought; several requirements such as double key entry into the vendor image tracking system and merging oversized documents into PDF files had not ever been done before by the vendor and had to be developed to work within their systems. Often these newly developed processes have had to be tweaked or reworked before they provided the desired output, adding to the schedule delay. In December 2010, the hearings files became accessible online. The original project was completed significantly under budget, with approximately \$93,700 remaining of grant funds. With grant funds expended to date on the East Texas project the Commission anticipates a cost of \$1.71 per image. In December 2010, the Commission also requested to expand the project to include Railroad Commission District 3 (Gulf Coast Region). The Commission estimates expanding the project will cost approximately \$80,000 to image and OCR additional files from Region 3 that will be added to existing database.

A website detailing progress on the *Historic Oil and Gas Hearings Files, 1932–1972* project may be found on the Railroad Commission's web site. The website provides a useful overview of the project's success to date, and is available at:

<http://www.rrc.state.tx.us/data/online/HistoricalHearingsHighlights.php>.

Summary of Project Activities July to December 2010

Quality Assurance (Task 2)

Neubus developed an editing tool that was implemented in February. This tool allows RRC staff to perform in-house page rotation, relocation, and deletion within a PDF file. All file editing and rescanning work was completed in October.

The individual hearings files imaged for this project were converted to PDF files from the TIFF format for the user interface. To accommodate files that are quite large in size, Neubus developed a new public user interface for this project that allows accessibility to the file pages without having to download the PDF to the desktop. The user interface was completed and the files were made available online on December 2, 2010.

Re-purpose Finding Aid (Task 4)

In December the Commission met with the Texas State Library and Archives Commission's Texas Heritage Digitization Initiative coordinator and the Neubus technical staff to discuss the development of a finding aid for the collection, and specifically the metadata and application program interface (API) needs of both the Railroad Commission and the Texas Heritage Online site. Neubus developed the API and provided the metadata at the end of December. The Railroad Commission designed a draft finding aid that is consistent with those used at the Texas State Library and Archives Commission and scheduled a meeting with the staff Archivist for the end of January to review the draft. It is anticipated that the finding aid will be available on the RRC web site by the end of February.

Promote Collections (Task 5)

During this time period the Commission continued to work with Neubus to develop a survey that appears on the historical hearings web site. Currently the survey does not function as envisioned; however Neubus is aware of the issue and is working to fix it. The survey collects data on the user's purpose for using the site. Currently the user clicks on a link to participate in the survey; in the future the survey will pop up and the user will be able to choose whether or not to participate. The survey results will be used to create a more comprehensive usage survey at a later date.

Neubus developed a method for measuring usage by counting the number of hits to the historical hearings query application separate from the other Railroad Commission image query pages they maintain. In December 2010, the first month of the web site, there were 1,676 hits. This usage compares with a total of 49 paper files accessed between September 2008 and September 2009 when all files had been pulled from the shelf for imaging and were therefore unavailable for use in that format.

Other promotional efforts during this period included a presentation and breakout session on the project at the Museum Computer Network conference in October and two workshops about the project, including hands on instruction of how to search the hearings files online at the RRC Oil and Gas Expo in November.

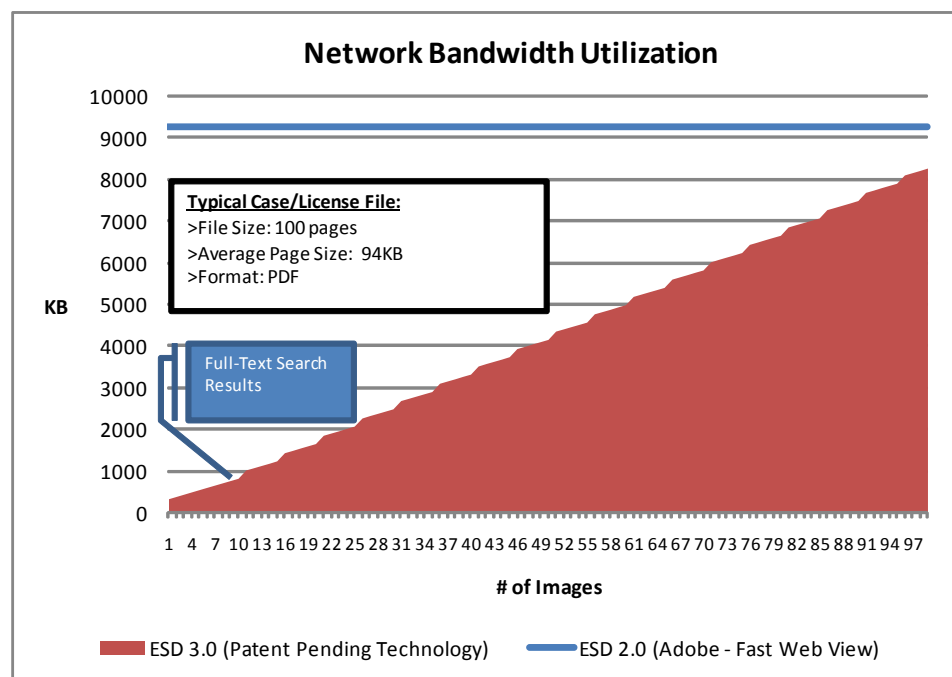
Project Assessment to Date

As previously reported, the Railroad Commission anticipated a more aggressive timeline than was practicable. Issues emerged as the contractor implemented processes required by the nature of the grant or the documents being digitized were not typical of their day-to-day work. The Commission and Neubus worked together to identify appropriate solutions before moving

forward, but working through issues was more time consuming than anticipated. The Commission is pleased with the high quality of the end result.

Scanned Documents

Neubus used a Bowe Bell and Howell Ngenuity scanner to scan the regular sized paper and a Contex SD4490 large format scanner for all documents over 11 inches x 17 inches. For the best image quality, the contractor ultimately provided all images in color, which increased file size tremendously. The Commission expressed concern about end users difficulty downloading and opening such large files. Neubus recognized the problem as well and developed a solution for this issue.



The new intuitive user interface addresses this problem by throttling the “intelligent delivery” of content, allowing for the display of image files of virtually unlimited length. The “intelligent delivery” is accomplished by enabling full-text search of the entire file or selective page viewing without requiring the entire file to be resident on the user’s desktop, as contrasted with a more traditional “Fast Web View” implementation, saving valuable network bandwidth resources for other uses. In addition, the bit-depth of the images displayed can be changed in real-time (reducing the image size) to accommodate extremely low bandwidth requirements.

Oversize Documents (Merge)

To maintain the integrity of each hearing file, Neubus developed a process for merging oversized documents into their correct place within a hearing file PDF. The QA and editing process revealed that not all oversized documents merged correctly, and some did not merge at all. When the files were reprocessed, some oversized documents merged correctly, but other issues such as duplicate documents, deleted pages on either side of the merged document, and pages generally out of order emerged. Since the reprocessing occurred after the initial QA was complete, RRC staff had to re-examine most files during the editing process and identified many pages that had

to be rescanned due to accidental deletion. Also, after reprocessing many merge oversize and true oversize bar codes appeared in the PDF files and had to be deleted. All files that include true oversized or merge oversized documents were checked again for quality after rescanning to make sure that all pages are in the file and in the correct order.

Editing Tool

Neubus developed an editing tool that allows RRC staff to perform in-house page rotation, relocation, and deletion within a PDF file. The tool is a web-based application and does not require additional software to use. The tool was used to correct the majority of the processing errors that occurred during the scanning phase of the project. It provided the ability to view each PDF file as it will be seen online as well as a thumbnail mode for editing purposes. Unfortunately, opening the files through the editing tool was a very slow process due to large file size and smaller bandwidth, which contributed to project delays. Oversized merged documents that could be moved from the PDF to become separate true oversized documents without file integrity degradation were identified during the editing process and were moved. This helped lower file sizes in a good number of files and ultimately allowed for faster accessibility times online.

Rescanning

Neubus developed a process for rescanning pages and adding them into the appropriate PDF files. The process placed the rescanned pages at the beginning of the appropriate PDF so that the editing tool could then be used to move the pages to their proper location within the file. Initial testing of the process determined that additional development work was necessary, as some documents could not be relocated from the beginning of the file while other documents had duplicate page numbers in the thumbnail edit mode and could not be moved independently of each other. The true oversized TIFF images could not be viewed at all. Neubus developed a solution to the issues. All of the necessary pages were rescanned and placed in the correct order within the appropriate file.

Imaging Cost

The hearings files became accessible online at the beginning of December. The original project has been completed significantly under budget, with approximately \$93,700 remaining of grant funds. With grant funds expended to date on the East Texas project the Commission anticipates a cost of \$1.71 per image.

Project Impact on the Grant-Receiving Institution

The Railroad Commission presented an overview of and findings from this project at the Museum Computer Network 2010 conference held October. Training sessions on the Historic Hearings query application were held at the Railroad Commission's Texas Oil and Gas Expo in November 2010. The Commission will collaborate with the Texas Heritage Digitization Initiative (THDI) for a presentation at the Texas Library Association annual conference in 2011 as well as develop a brochure for the THDI coordinator to disseminate at a booth at the Texas Historical Association conference in March 2011. In December 2010, the Commission requested and received approval for a 6-month extension of the project to better measure the success through both usage statistics and a survey.