

NHPRC Grant NAR09-RD-10019-09
Digitizing Historic Oil and Gas Hearings Files, 1932–1972
Semi-Annual Narrative Progress Report, July 31, 2010

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 the second year all of the 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 has taken much longer than originally anticipated, delaying the schedule by approximately six months. The project has proved to be significantly more complex than initially anticipated. The Railroad Commission's project was the first project of this nature for the vendor, and as such a period of trial and error generally followed as each new procedure was implemented, further delaying the initial timeline. As an example, this was the first use of double key entry into the vendor image tracking system and merging oversized documents into PDF files. The Commission is pleased with the high quality end results.

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/imagedrecords.php#hearings>.

Summary of Project Activities January to June

Scanning Process (Task 3)

During the prepping process, the files were divided into three categories for scanning. The categories were determined by type and purpose of each document. The categories are:

1. Base or Parent Boxes: For each hearing file these contained all paper that was ledger-sized (11 inches x 17 inches) or smaller. The majority of the paper in each file was included in this category. There are 130 Base or Parent boxes and all were scanned by December 2009.
2. Merge Oversize Boxes: The hearings files contain many large maps and other exhibits. It is important for all exhibit documents in the imaged file to remain in the same order as the paper was in the original file. Oversized items have to be scanned on a large format scanner as separate files and Neubus generally treats them separately, providing access to

the images with links adjacent to the link to the base or parent file PDF. For this project, it was necessary to merge the images of the oversized exhibit documents into the PDF file of the parent file so that all images remained in the correct order. There are 36 Merge Oversize boxes. Scanning of these documents was completed in January 2010.

3. True Oversize Boxes: Some oversized documents such as well logs could not be merged into the PDF due to their size. These were also scanned on the large format scanner but were treated as separate files. A link to each image will appear adjacent to the link to the parent PDF file in the display for each hearing file. For each of these oversized items a descriptive part of the paper document was copied during the preparatory process and placed in its correct place in the parent file with a note that an image of the complete document is available as an Oversized Document link. There are 43 True Oversize boxes. Scanning of these documents was completed in February, 2010.

Quality Assurance (Task 2)

Initial quality assurance work (QA) on all boxes was completed in April, 2010. Neubus developed an editing tool that allows RRC staff to perform in-house page rotation, relocation, and deletion within a PDF file. This tool streamlined the QA process, and enhanced the Commission's control over the editing process. File editing is approximately 85% complete.

Re-purpose Finding Aid (Task 4)

In May the Commission met with the staff archivist at the Texas State Library and Archives Commission to discuss the development of a finding aid to the collection. The finding aid will be consistent with the online finding aids of the Texas State Archives and the staff archivist will provide information and assistance to ensure it meets their standards.

Promote Collections (Task 5)

During this period the Commission worked with Neubus to develop a survey that will appear on the web site each time a user initiates a query session. The survey will collect data on the status of the user and the purpose for using the site. The user will be able to choose whether or not to participate in the survey. The initial survey results will be used to create a more comprehensive usage survey at a later date.

Neubus is developing 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.

Project Assessment to Date

The Railroad Commission anticipated a more aggressive timeline than was practicable. The Commission sought to address as many concerns and potential issues as possible through the statement of work process to avoid problems once digitization began, but each task of the project has taken much longer than anticipated to complete. Throughout each phase of the project, issues emerge as the contractor implements processes required by the nature of the grant or the documents to be digitized that are not typical of their day-to-day work. The Commission and Neubus continue to work together to identify appropriate solutions before moving forward, but working through these issues is more time consuming than anticipated, resulting in an

approximately six month delay to the project. As the QA and editing processes reach completion, 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 provided all images in color, which greatly increased file size. The Commission expressed concern about difficulties end users may encounter downloading and opening such large files. Neubus is actively seeking a solution to this concern.

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 emerged such as duplicate documents, deleted pages on either side of the merged document, or pages that were out of order. Since the reprocessing occurred after the initial QA was complete, RRC staff found it necessary to re-examine most files during the editing process and identified many pages that will have to be rescanned due to inadvertent deletion. Additionally, after reprocessing many Merge Oversize and True Oversize bar codes appeared in the PDF files. All files that include true oversized or merge oversized documents will be assessed again so that the bar code pages can be deleted.

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. The tool is used to correct the majority of the processing errors that occurred during the scanning phase of the project. It provides the ability to view each PDF file as it will be seen online, as well as a thumbnail image for editing purposes. Unfortunately, large file size and smaller bandwidth contribute to project delays, as opening the files through the editing tool is a very slow process. Oversized merged documents that can be moved from the PDF to become separate true oversized documents without file integrity degradation are identified during the editing process and moved. This helps decrease file sizes in many files and will allow for faster download times for end users.

Rescanning

Neubus developed a process for rescanning pages and adding them into the appropriate PDF files. A test was conducted to assess regular pages, as well as merge and true oversized documents. The process places the rescanned pages at the beginning of the appropriate PDF so that the editing tool can then be used to move the pages to their proper location within the file. The test 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 is seeking a solution to these issues. Once the process is working correctly, it is anticipated that the rescan process will be completed within a month.

Project Impact on the Grant-Receiving Institution

The Railroad Commission presented an overview of this project at the Texas Heritage Digitization Initiative annual conference in February 2010. The presentation focused on the Commission's experience using an outside vendor to complete an archival imaging project, as well as the grant application and funding process with the NHPRC. The Commission recently learned that its proposal to present at the Museum Computer Network 2010 conference in October was accepted. The conference is a unique opportunity to present findings from this project, as a relevant resource that will be readily available for background research to develop exhibit material for curators and museum staff of local and state history museums; museums developing exhibits related to oil and gas, energy, industrial, political, and economic regulation topics; and exhibits portraying mid-twentieth century small town and rural life and work. Training sessions on the Historic Hearings query application will be held at the Texas Oil and Gas Expo in November 2010. There are also plans to collaborate with the Texas Heritage Digitization Initiative for a presentation at the Texas Library Association annual conference in 2011.

The Railroad Commission anticipates that a no-cost extension to the grant period may be necessary to acquire more complete data about increase in use, and any changes to user demographics as anticipated in the initial grant application. Additionally, the Railroad Commission anticipates that the original project will be completed significantly under budget and would like to expand the project to include Railroad Commission District 3 (Gulf Coast Region). The Commission estimates expanding the project will cost approximately \$65,000 to image and OCR additional files from Region 3 that will be added to existing database. The Commission will deobligate these funds at the end of the grant period if an expansion of the project is not approved. The Commission requests a no-cost extension of six months to complete the analysis phase of the project, or a yearlong extension if expansion of the project is approved.

Expansion of the project to include District 3 will make available additional documents with both historic and contemporary value to a wider audience, and have the potential to influence current discourse regarding drilling policy in the Gulf of Mexico. This district includes the site of the original commercial oil production in Texas at Spindletop, which was discovered in 1901 and led to an oil exploration and production “boom” in Texas. It prompted a transition from rural farm life to urban industrial society as Americans began to work in the oil fields and benefit from the development of a petroleum-based economy. This district also includes the Texas Gulf coast where modern day drilling techniques, petroleum geology methodology, and refining processes were developed in the early 20th century. Exploration and production in bay and offshore waters began in the 1930s. The Commission anticipates it could complete the hearings files from the additional district with the funds that will remain upon completion of the original three district (Districts 5, 6, and 6E) project. The proposed expanded project would digitize approximately 14,500 additional files, that contain 275,000 base, or regular sized paper, images, and 20,000 oversized images.