

NHPRC Grant NAR09-RD-10019-09
Digitizing Historic Oil and Gas Hearings Files, 1932–1972
Semi-Annual Narrative Progress Report, January 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 six months of the grant period. The Commission anticipated that Task 2, Prepare Materials for Digitization, would be complete, and that much of activity during the period would focus on repurposing the finding aid and promoting the collection. As reported in July, the Planning and Development phase took longer than anticipated in the project proposal, delaying the schedule by approximately three months.

The initial delay postponed scanning from the originally scheduled completion date of September 2009, but as of December 2009 scanning is complete for all base box materials and is approximately 75 percent complete for the oversized documents. 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 July to December

Project Specifications (Task 1)

- The Commission did not finalize and approve the Project Specifications document until September 8, 2009. The document provides more detail about the work to be completed and results to be achieved as outlined in the Statement of Work. Following the approval of the Project Specifications, a limited number of boxes were scanned for initial review before moving forward with the entire collection.
- The Commission anticipated that Phase I would be complete in three months, rather following eight months of planning and development work the imaging process began. Additional development will be needed as methods for tracking web usage are designed and a survey for determining successfulness of outreach is created.

Preparing Physical Files for Imaging (Task 2)

- Concurrent with the development of the detailed project specifications, preparation of materials continued throughout the summer months and was completed at the beginning of September.
- Neubus, Inc. provided 2.5 temporary employees to prepare 125 boxes of files.
- The contents of each file were placed in the correct order, with special attention paid to fragile or unique pages that required special handling. Oversized documents that would need to be scanned on a large format scanner were identified, while the staff also decided if each oversized item would be merged into the main file once scanned and converted to a PDF file or if the document would remain as a *stand-alone* oversized document.

Key Entry Process (Task 2)

- Neubus, Inc. uses a bar code scanning system to track items through the digitization process. To meet Neubus' requirements, the key fields identified from the file folder label on each file had to be key entered into their web-based Production Control System (PCS) application. Key Entry began in July and was completed at the end of September.
- The key field database created by the PCS tracking system will be repurposed as a finding aid when scanning is complete.

Scanning Process (Task 3)

Scanning began in July with three initial boxes available for review before proceeding with the remainder of the project. 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 nearly complete at the end of the reporting period.
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. At the conclusion of the reporting period approximately half of these boxes are scanned.

Quality Assurance (Task 2)

- Quality assurance (QA) work was done on the initial three boxes between September and November.
- QA activities on the remaining scanned files began in late December.
- Neubus provides a private test web site to review the hearings files. Upon completion of the QA process and file approval, the files will move to the public web site, accessible through the RRC web site.

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. As reported in the July 2009 Semi-Annual Narrative Report generally issues emerged when the contractor was required to implement a process required by the nature of the grant or the documents to be digitized that is not typical of their day-to-day work. In each situation the Commission and Neubus were able to work together and identify an appropriate solution before moving forward, but working through such issues was a more time consuming process than initially anticipated. Initial reviews through the QA process indicate that the end result will be of high quality.

Double Key Entry

The issues related to the double key entry process that were reported in the July 2009 report continued into July and August as Neubus reprogrammed their PCS system and determined how to best use PCS to associate the “merged oversized” and “true oversized” documents with the base or parent part for each hearing file. As the Neubus PCS system is not only used for tracking documents, but also accurate processing of scanned images as well as indexing, the three categories of documents identified for this project brought a level of complexity that required the vendor to develop a unique process to manage. Ultimately, the solution they developed worked well and will be used by the Commission in the future.

Oversized documents (Merge)

One critical factor throughout this project is maintaining the intact integrity of each hearing file. The ability to merge images of oversized documents into their correct place within a hearing file PDF is an important requirement. Neubus developed a bar code system that works well for this purpose. The PCS system generated two identical bar codes when an oversized document that required merging was key entered into the system. One bar code was attached to the document for scanning and the other was placed in the parent file as a placeholder or marker. During post-scanning processing the place marker bar code and document bar code were matched up and the oversized document was merged during PDF conversion.

Adding oversized documents to a PDF file has, in some instances, created very large files. To accommodate users a message with the option to download the file instead of opening it will accompany all PDF files with oversized documents on the web site. The message states:

“This PDF document contains oversized pages and may take several minutes to open. Click [\(download\)](#) to download without opening PDF.”

Scanning

The varied quality of the original documents made obtaining good quality scans a challenge throughout this project. Paper ranged from thick to thin and included types as varied as newspaper print copies, microfilm copies made with the reigning process of the time, carbon copies, telegrams, and “negative” copies. To best scan the varied paper types and qualities Neubus purchased a new model Bowe Bell and Howell Ngenuity scanner for this project. It was difficult to set the scanner settings with such variation in original quality; however after much testing and consultation with the scanner software vendor, they were able to adjust the settings to achieve the best digital images possible.

Achieving good quality images from the “negative” copies proved to be the biggest challenge. To address this challenge, Neubus developed three “profiles” that the scanning technician can choose from for imaging “negative” copies, one for dark, medium, or light black backgrounds. With the image quality from “negative” copies greatly improved, one of the post-scanning processes converted the negative images to positive images, with varying results. The Commission chose to leave the images as negatives to accurately reflect the paper files. Neubus corrected this so that all of the images are now high quality.

Initial Three Box Process

Scanning and the initial three-box review process occurred concurrently for several months. Initially one box of hearings files was scanned and the respective oversized documents merged and converted into a PDF file for an initial review. The box included the files for eight hearings. Image quality issues were immediately identified with certain types of paper, such as the “negative” copies. Mistakes made during the preparatory phase of the project were also identified, which led to discussions of how to fix the files without having to rescan the files. At this time two additional boxes, for a total of 64 files, were readied for initial scanning and review to provide a larger sample size. The initial review led to discussions with Neubus and adjustments in scanning settings and processes, with the end result being high quality images.

Project Impact on the Grant-Receiving Institution

The Railroad Commission will present findings from this project at the Texas Heritage Digitization Initiative annual conference in February 2010. The Commission will submit a proposal to the Texas Library Association for presentation at their annual conference in 2011. The Commission is also pursuing other options to present its findings from this project in 2010, however, such opportunities may be limited by anticipated budget reductions related to ongoing fiscal uncertainties facing the State of Texas.

The Texas Legislative Council visited the Commission after learning of this project to discuss scanning techniques and the technology employed for this project.