

**NHPRC Grant NAR09-RD-10019-09**  
**Digitizing Historic Oil and Gas Hearings Files, 1932–1972**  
**Final Narrative Progress Report, September 30, 2011**

**Performance Objectives:**

- 1) Timely submission of complete reports, which include detailed cost analyses for each part of the project, as well as three copies of grant products such as digitizing guidelines, publicity materials, and finding aids.**

All required narrative and financial reports were submitted to NHPRC to meet each deadline. No hard copy products were produced as a result of this grant.

- 2) Scan the approximately 120,000 pages from the East Texas Oil Field (Districts 5, 6, and 6E) hearings between 1932 and 1972.**

124,908 pages, including oversized plats, maps, engineering drawings, and well logs, were scanned. They are available online at <http://rrcsearch3.neubus.com/esd3-rrc/>

- 3) Make the scanned images available through three methods: links from the new finding aid, links from Texas Heritage Online, and a searchable East Texas Oil Field website to allow searching and browsing by the index fields.**

The images are available through the finding aid available online at <http://www.rrc.state.tx.us/data/online/findingaid.php> as well as through the searchable East Texas Historical Hearings Files web site at <http://rrcsearch3.neubus.com/esd3-rrc/>. The Texas Heritage Digitization Initiative, which produced the Texas Heritage Online site, was eliminated from the Texas State Library and Archives Commission as a result of budget reductions; therefore links to the images could not be made available through that resource.

- 4) Develop a website that publicizes the project and describes the processes and costs associated with preparing and scanning, and making available online these collections, especially discussing the ability of a scanning vendor to meet archival standards for handling documents.**

The Railroad Commission developed a project highlights web site to publicize and document the progress of the East Texas Historical Hearings project. The project highlights page is available online at <http://www.rrc.state.tx.us/data/online/HistoricalHearingsHighlights.php>. The page includes links to the Commission's semi-annual and annual narrative reports to the NHPRC, which detail the challenges encountered and results accomplished by working with a vendor for digitization.

- 5) Track and report on the project website about the usage of the collections prior and after digitizing in terms of user satisfaction, number of reference requests and usage of the originals.**

The monthly statistics for usage of the imaged collection are posted on the project highlights page available at:

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

- 6) Keep the costs of the project below approximately \$2.80 per scan and reduce the cost per use of a file below \$17.00.**

The total cost of this project, including funding from the NHPRC and the Railroad Commission's match, was \$224,666.15. 124,908 pages were imaged, equal to a cost per scan of \$1.80. As of August 31, 2011, the East Texas Historical Hearings web site had been used a total of 9,710 times, bringing the current cost per use of a file to \$23.14. At the current access rate, it is anticipated that the cost per use of a file will fall below \$17 in December 2011, a year after the files became available online.

- 7) Develop an outreach plan that results in a measurable increase of use of these records by historians, environmental scientists, economists and other researchers who rarely use the current files.**

Outreach activities included posting the availability of the new East Texas Oil and Gas Hearings images on the home page of the Railroad Commission web site as well as giving presentations about the new web site at conferences that drew diverse audiences. The Annual Texas Heritage Digitization Initiative conference was attended by a number of academic and public librarians from across the state. The Museum Computer Network conference attracted an international attendance of historians, museum curators, academicians, and researchers from North America and Europe. The Oil and Gas Expo drew an audience of engineers, geologists, and environmental scientists specifically interested in Texas oil and gas. The Texas Library Association conference included librarians from public and academic libraries as well as librarians from schools across Texas where these records will provide important research material.

A survey was also added to the East Texas Historical Hearings web site to determine the purposes for visits to the site. Through August 2011, 232 responses to the survey were received; the purposes for visiting the site were:

<b>Purpose of Visit</b>	<b>Number of Survey Respondents</b>
Business Case Analysis	8
Engineering Analysis	21
Economic Analysis	0
Geological Analysis	16
Higher Education	1
Historical Analysis	12

<b>Purpose of Visit</b>	<b>Number of Survey Respondents</b>
Legal Research	16
Other	18
Oil and Industry Supported Research	140
Secondary Education Research	0
Total	232

**8) Present on the project during at least one professional conference.**

Presentations at professional conferences included:

Texas Heritage Digitization Initiative annual conference, February 5, 2010  
Museum Computer Network annual conference, October 28, 2010  
Railroad Commission of Texas Oil and Gas Expo, multiple workshops given on November 4 and 5, 2010  
Texas Library Association annual conference, April 14, 2011

**Summary of Project Activities**

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.

The project was divided into three phases:

Phase I: Planning and Development  
Phase II: Prepare Materials and Digitize Materials  
Phase III: Dissemination, Initial Evaluation, and Sustainability

The project was 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 process had to be developed to work within their systems. Often these newly developed processes had to be tweaked or reworked before they provided the desired output, adding to the schedule delay. The contractor was always receptive to the Commission's concerns or issues, and appropriate resolutions were implemented before moving forward to the next phase of the project. The Commission was successful in digitizing 124,908 pages of oil and gas hearings records from the East Texas region of the state, covering the period 1932-1972, and in December 2010, the hearings files became accessible online. An online finding aid was also created to enhance access to the records. The project was completed significantly under budget, with approximately \$93,700 remaining of grant funds.

## Phase I: Planning and Development

Phase I of the project included the development of a Statement of Work and Project Specifications with the vendor outsourced to provide imaging services.

As detailed in the grant application, for this project the RRC used Neubus Inc., the outsourced imaging vendor for the State of Texas, to digitize project files. A Statement of Work (SOW) was prepared to outline the work to be done and results to be achieved. To understand the requirements of this project, Neubus met with representatives of the Railroad Commission, the Texas State Archives, and the Texas Heritage Digitization Initiative, which all had roles in the project.

Once the SOW was approved, the Project Specifications document was developed to detail the requirements as outlined in the SOW. Several issues were encountered during this step in the process that had to be resolved before the Project Specifications document could be approved and the project could move forward. These issues included maintaining file order and integrity with oversized items interspersed with regular sized paper and the ability to print oversized pages from within a PDF file. Following the approval of the Project Specifications document, a limited number of boxes were scanned for initial review before moving to Phase II. The Commission anticipated that Phase I would be complete in three months, however the planning and development work took nearly eight months.

## Phase II: Prepare Materials and Digitize Materials

Phase II of the project included preparing the physical files for scanning, key entry process, scanning the pages, checking each scanned page for quality, and rescanning as necessary.

The physical preparation work included placing the contents of each file in the correct order and paying special attention to fragile or unique pages that required special handling. Oversized documents, including maps and other large exhibits, 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. Neubus provided temporary employees to assist with document preparation of 125 boxes of files.

In the process of prepping the files for scanning, project staff discovered that the exhibits submitted in support for several hearings are pages from a copyrighted loose-leaf publication of *Special Field Rules for Texas*, published by R.W. Byram & Co. The Commission contacted R. W. Byram and received permission to image the *Field Rules* copies contained in the files, as long as each exhibit is limited in scope. Some of the hearings files also contain newspaper pages of legal notices that were required to be published before a hearing was held. When an entire newspaper page was included in the hearing file, the appropriate legal notice was copied for scanning so that newspaper copyright would not be violated.

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. Neubus decided that a double data entry process would be best for this project; however reprogramming of their web-based application was necessary to accept double data entry. Additionally, it was necessary to determine how to best use PCS to associate the “merged oversized” and “true oversized” documents, each in separate boxes for scanning, with the contents of the base (or “parent”) box and continue to track all documents as they moved through the imaging process. It took several months for the issues of double data entry and oversized document tracking to be resolved satisfactorily so that Neubus did not require any more additional work than necessary for tracking purposes. The key field database created by the PCS tracking system was later repurposed as a finding aid when scanning was complete.

Scanning was accomplished using two types of equipment: one for regular sized paper up to 11 inches by 17 inches and a large format scanner for documents larger than what could be scanned on the regular machine. During the prepping process, the files were divided into three categories for scanning. The categories were determined by type and purpose of each document and included base or parent boxes for regular sized paper, merge oversized boxes for oversized documents that would be merged back into the base file to maintain the correct order, and true oversized boxes that included documents such as well logs that could not be merged back into the base file.

Once scanning was complete, quality assurance (QA) work was done to assess the quality of the scanned images. Neubus developed a web-based editing tool that allowed 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. 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 had 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 had to be assessed again so that the bar code pages could be deleted.

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 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. 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 was able to develop a solution for these issues; but due to these issues completion of the QA tasks took longer than originally anticipated.

### Phase III: Dissemination, Initial Evaluation, and Sustainability

Phase III of the project included providing access to the scanned images from the RRC web site, creating a finding aid to the hearings files, promoting the collection, and creating links to the files from the Texas Heritage Online portal.

The East Texas Historical Hearings files became available on the RRC web site in December 2010. 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.

The Commission worked with its assigned State Archivist at the Texas State Library and Archives Commission to create an online finding aid to the digitized collection that is consistent with the finding aids available through the State Archives. Neubus provided the metadata and application program interface needed for the finding aid. The aid is available at: <http://www.rrc.state.tx.us/data/online/findingaid.php> .

The Commission promoted the online collection in several ways. Staff worked with Neubus to develop a survey that appears on the historical hearings web site. The survey pops up at the beginning of each research session and collects data on the user’s purpose for using the site. The results of this survey are reported in the performance objectives section of this report.

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. From December 2010 through August 2011 there have been 9,710 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. The monthly statistics for usage of the imaged collection are posted on the project highlights page available at <http://www.rrc.state.tx.us/data/online/HistoricalHearingsHighlights.php>.

Other promotional efforts included presentations at four different conferences to audiences with varied interests and backgrounds. More details about the diversity of conferences attend is reported in the performance objectives section of this report.

The Commission was unable to fulfill the performance objective related to the Texas Heritage Digitization Initiative and its Texas Heritage Online portal. Legislatively mandated budget reductions at the Texas State Library and Archives eliminated the program.

### **Project Impact on the Grant-Receiving Institution**

The Railroad Commission has taken the lessons learned from completing this project and applied them to a new project, funded by the Commission, to digitize all newly closed oil and gas hearings and make them available online. This will allow for greater access to hearings files in a timely manner as well as provide efficiencies for the records staff. In addition, all existing Commission online imaging projects are being migrated to the new user interface developed for the grant project as it provides a faster, more efficient, way to access all files.