

2002

The Conversion of the Mississippi State Sovereignty Commission Records

Anna Schwind

Mississippi Department of Archives and History

Sarah Rowe-Sims

Mississippi Department of Archives and History

David Pilcher

Mississippi Department of Archives and History

Follow this and additional works at: <http://aquila.usm.edu/theprimarysource>

Recommended Citation

Schwind, Anna; Rowe-Sims, Sarah; and Pilcher, David (2002) "The Conversion of the Mississippi State Sovereignty Commission Records," *The Primary Source*: Vol. 24: Iss. 2, Article 2.

DOI: 10.18785/ps.2402.02

Available at: <http://aquila.usm.edu/theprimarysource/vol24/iss2/2>

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in The Primary Source by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.

The Conversion of the Mississippi State Sovereignty Commission Records

by Anna Schwind, Sarah Rowe-Sims and David Pilcher

In 2001, after the final court-ordered record release, the Electronic Records (ER) section of the Archives and Library Division of the Mississippi Department of Archives and History (MDAH) assumed responsibility for the electronic version of the Mississippi State Sovereignty Commission records.⁷ In addition to basic maintenance and preservation responsibilities, ER was assigned the task of converting the proprietary in-house system to make it web accessible. This was a watershed project representing both closure and new beginnings. Not only would the web-enabled electronic version mark the final stage in access with the promise of global availability via the Internet but it was also the first time ER would be able to test its migration strategies and open source philosophy.

Although in technological terms the year 1994 is ancient history, the scanning project was no less daunting or ambitious than the recent conversion. The original Commission records were scanned, indexed and processed by MDAH in accordance with American Civil Liberties Union v. Fordice, 969 F.Supp. 403 (S.D.Miss. 1994). MDAH needed to be able to meet the court's special privacy and disclosure provisions while preserving the original hardcopy records and maintaining the integrity of the original filing system. MDAH sought an imaging solution that would enable the production of an exact and secure electronic version of originals in which every item was uniquely identified and that mirrored the original filing system. These images would be coupled with and accessible via a court-required index of personal names. In addition, MDAH needed the ability to affix court-specified redactions to portions of text on certain pages as well as seal folders of documents upon request.

⁷ For those unfamiliar with the Mississippi State Sovereignty Commission, the protracted legal battle to open its files and MDAH's involvement in the whole saga, see the Agency History and Case Chronology of Sovereignty Commission online
<http://mdah.state.ms.us/arlib/contents/er/scagencycasehistory.html>

MDAH contracted with Alabama based SysCon Computers to design and implement an imaging system. The project staff worked with SysCon to theorize how input, retrieval and access should work within the database. SysCon proposed to develop and install in three phases a custom application for project use. The first phase would address the scanning, storage and partial indexing of documents. The second phase would address the complete indexing and redaction of scanned documents. The third phase would provide for search, retrieval and printing of pages. SysCon delivered and installed the hardware and initial software in November 1994. The database, which ran on Oracle and used a Windows based client, was customized for project purposes by SysCon President Bob Wilson and was modified and upgraded as necessary by him throughout the project.

The system included a separate form for quality control viewing of images as they were scanned. Pages were rescanned as necessary. As images were created they were stored on the network drive until sufficient volume of 655 megabytes was available to create a compact disc. The images were then exported to a single data file and one index file, a process known as BLOBbing.⁸ These two files were then transferred to digital audiotape and copied to a compact disc using a recording unit. A second disc was made for backup. The contents of the disc were then compared with the images still stored on the network drive. If they compared exactly, the image files were removed from the network drive.

Eventually twelve compact discs were filled with image data files. However, later in the project problems arose with discs jamming when retrieving images sequentially from separate discs within the same changer, causing work delays. Therefore it became necessary to transfer the images on the discs back onto two 4.3-gigabyte hard drives which SysCon installed on the server. This greatly increased speed and responsiveness during the period of searching and printing following the inquiry period. A writable compact disc drive was added to one of the workstations at this time. It was used first in restoring information to the hard drive and later to produce compact discs for the public access version of the system.

Scanning began in early December 1994 and was completed with the exception of the Commission's financial records in late May 1995. The financial records were scanned several months later, bringing the total of images within the database to approximately 132,700. On average, roughly 1000 images were scanned and checked per day. The indexing of the scanned images was done within a separate form which allowed the calling up of the image and name entry forms side by side onscreen and the linking of personal names and classification (as Victim or State Actor) with each image. Once indexed, images were inaccessible from this form as they were now part of the DBMS BLOB. The quality control form could then be used to check for indexing errors, with a separate form available as well for making corrections. Eventually all three project staff members were engaged in indexing full-time, with additional staff members brought in to assist with indexing quality control. The indexing and quality control period lasted approximately one year and two months. This resulted in a personal name index of approximately 87,000 unique names, comprising nearly 300,000 occurrences, with each name classified as a Victim or State Actor.

In January 1997, in accordance with the court order, MDAH advertised in local and national newspapers inviting the public to review the records. MDAH mailed questionnaires to the nearly 1000 people who responded to the ads. 700 completed questionnaires were returned within the specified time. Based on the identifying information provided in the questionnaires, nearly 360 of these individuals were located in the files and mailed copies of the salient documents. This mailing amounted to roughly 25,000 pages.

⁸ BLOB: "short for *Binary Large Object*, is a collection of data stored as a single entity in a database management system (DBMS). BLOBS are used primarily to hold multimedia objects such as images, videos, and sound, although they can also be used to store programs or even fragments of code. Not all DBMSs support BLOBs." ["BLOB," Webopedia.com, September 01, 1997<

<http://webopedia.com/TERM/B/BLOB.html>> (January 2, 2003)].

The court required that MDAH redact the names of other respondents from the printed documents mailed to respondents. The system had the ability to attach an unlimited number of rectangular overlays to a name and to activate and deactivate as needed to redact the name and identifying information. When activated, the redactions appear as black boxes on the image. Upon reviewing their records, the majority of individuals chose full disclosure. However, 42 individuals chose some kind of privacy option requiring adjudication.

In January 1998, Judge Barbour ruled that all non-contested records should be opened and in preparation, the project staff created a new set of compact discs with each block of data run through a purge program to remove the contested images completely. Approximately 7700 pages remained sealed and names of contested individuals were removed from the index. On March 17, 1998 the bulk of the records were opened. A separate server, database and computer terminals were set up in the search room for this purpose. Subsequent releases occurred on July 31, 2000 and January 18, 2001. These releases contained the court-approved redactions and approximately 2,311 pages of rebuttal material submitted by Victims in response to the records. In accordance with the court's directive, the rebuttal material was scanned in as an accretion to the main collection. Statements were added to every page in which the name of a rebuttal respondent appeared alerting the user to the existence of his/her rebuttal. The records were searchable by personal name, folder title, and classification plus folder number. These capabilities grew from access requirements during processing.

The 2001 release marked the end of the privacy and disclosure process and ER assumed responsibility for the electronic version of the records. With the short-term court ordered objectives completed, ER now turned its attention to the long-term requirements of preservation and access. One of the underlying tenets of electronic records preservation is active intervention. Unlike their more stable hardcopy counterparts, hardware, software and file formats all have a tendency to display remarkably short life spans, and to be so dependent on one another that preservation of the data embedded in these systems invariably fails without periodic migrations. Therefore, a commitment to preserving electronic records is essentially a commitment to migrate them from platform to platform forever and ever. For ER the Sovereignty Commission Records provided the first opportunity to test our hand at the theory of constant migration for record preservation. The database had been built on a version of the Oracle database that was no longer supported by the company. The networking used to communicate between the database and the stations in the search room was an obsolete version of Novell, also unsupported, and without any way to integrate these stations into our intranet or the wider Internet. Lastly, the hardware was between six and eight years old and had already outlived its expected lifetime. The system, its database, and the records themselves were all living on borrowed time.

Rather than waiting until the staff arrived one day to find everything no longer worked, ER began the work of migrating the records to another platform and another system. According to our strategies for managing electronic records, we selected open source software whenever it was available to suit our needs. We selected open formats whenever available as well. In addition, our desire to provide increased access to the records put some other constraints on our migration efforts. We wanted to provide universal, worldwide access to the Sovereignty Commission images, which meant making them available through the World Wide Web. This limited the range of our choices for image formats, as we felt very strongly that the images should display natively in the browser, without requiring the use of proprietary plug-ins.

The Sovereignty Commission files were originally released in various iterations, because of court rulings and the privacy issues related to the people named in the files so the data was handled in various different ways in the database. In some cases the images were blobbed into the database with pointers as to their length and offset. In other cases the images were not blobbed, and just residing on the hard drive or the cd-roms, with pointers to their path and filename. As for the database indices, the indexing requirements changed several times during the course of the data entry, so that in some cases tables held duplicate data, while in other cases tables were empty of any rows, and in still others data that was of the same type and should have been in the same table was scattered across various parts of the database. Now, with the data in a fixed form, it became part of the migration task to clean up the database. It was even hoped

that some things that had been requested of the vendor for the original system but unavailable due to the technological limitations of the original system might be implemented in the migration, providing not only wider access, but also deeper access to the records.

The migration process involved a number of steps and took over a year to complete.

The first step was the purchase of new hardware.

The second step involved choosing the software. All the tools chosen for the project were open source and freely available: *linux*, *postgres*, *apache* and *php*. Since we wanted the images to ultimately appear on the web, we decided that the browser would be the access tool for both the in-house and the external version instead of trying to do a lot of effort duplication to create two versions. This choice automatically set even more constraints on us, including forcing us to change the look and feel of what had been a windows oriented interface to a more platform independent web oriented interface. We underestimated the impact this superficial change would have on the staff that used it on a daily basis.

The third step was the most processor intensive and time consuming step, although it typifies the type of things that computers are best suited to and involved next to no effort from the staff. The images that had been integrated into the database (the vast majority of them) had to be extracted from the database and made readable by any program. This required thousands of hours of processing time on our new hardware.

The fourth step involved converting the images from their original format to a new format. The original format of the images was a compressed bitonal tiff that would not display in any browser natively. ER considered various image formats in terms of image clarity, open formats and image sizes and had begun a conversion to *jpgs* before settling on the *png* format.

The fifth step involved extracting all the data, cleaning it up and reconstructing the database in *postgres*.

The sixth step was the one that involved the most hands-on work from the Electronic Records Section. In order to access the records we designed a user interface with various query capabilities. Besides adding extra capabilities to the system, such as the ability to browse the index and the ability to go directly to rebuttal files, we also wanted to make sure we included all the functionality of the old system. Not only that, we wanted the interface to be very easy to use. The creation of the interface also coincided with the data check phase. As queries were made available to ER staff who had worked on the previous version and knew the data fairly well, we were able to run queries and examine result sets and individual images to make sure everything was in order thus far. We were confident and hopeful that a leaner, cleaner database and new hardware would give us a significant performance increase over the aging original system, and it did. Queries returned results substantially faster than in the old version and queries that returned result sets that were too large for the old version to handle worked in the new version without crashing.

The new Sovereignty Commission database was ready to be opened, and it was decided that it would first be opened in-house with the hope of receiving feedback from staff and patrons and fixing any bugs we might discover in daily use of it before making it accessible on the Internet. In the first few days of use several critiques were made, including complaints about the clarity and size of the converted images. All of our testing had not focused on the images themselves, but on the search mechanism and we realized that we weren't exactly well suited to determining what image size would be appropriate. We assembled a committee to assist us in determining what resolution and size of image would be adequate and knew that we would soon be faced with a reconversion of all the images. We felt that soliciting input from staff members from different walks within MDAH, viewing this from different angles, would help us to identify all the necessary improvements. A new size and resolution were chosen, and the images underwent an

additional reconversion. The database stayed up and accessible throughout this process, and the new images replaced the old as they were converted.

As the images underwent closer scrutiny the issue of photographs was re-explored. Although the vast majority of the collection is paper in the form of reports, correspondence and newspaper clippings, there are approximately nine hundred photographs. These remarkable photographs include Freedom Rider and other activist mug shots from around the state and South, demonstrations in Mississippi, Alabama and the California-Berkeley campus and public relations photographs of African-American professionals, businesses, church activities and educational facilities. Access to these images has been poor. Pursuant to the court order records were indexed for personal names and thus a researcher would only *stumble* across a photograph if a name happened to appear and was indexed on that page, which is an uncommon occurrence. The project staff had long been unhappy with the status of photographs in the collection and wanted at the very least to provide an additional photograph index. In addition, ER also saw the need to integrate these images with the emerging MDAH digital image archive to provide seamless access to the agency's holdings. Such an endeavor would require the photograph scans to conform to the digital image standards laid out by the Archives & Library Division's Image and Sound Section (IS). Accordingly, ER scanned all the unredacted photographs and individually described each according to the ER/IS metadata element set. As of this writing, ER is in the process of constructing a search mechanism that will index them separately as well as link them back to their original folder in the main electronic collection. We expect this exciting new resource to be available as part of Sovereignty Commission Online in spring 2003.

With the second image conversion concluded, ER was ready to publish the Sovereignty Commission Online (as the new database was named) to the web. Those records, the first of many, many migrations which will hopefully provide increasing access to collections of electronic records accessioned by MDAH, are currently available at this URL : <http://www.mdah.state.ms.us/arlib/contents/er/> . We welcome any questions and feedback on this exciting project.

Please address any email about the project to elecsecs@mdah.state.ms.us

(Sarah Rowe-Sims has worked as an archivist in the MDAH Electronic Records Section since February 2001. She joined MDAH in 1994 and was one of a team of three archivists who processed and handled the opening of the Mississippi State Sovereignty Commission records from 1994-2001.

Anna Schwind has been the Lead Systems Administrator for the MDAH Electronic Records Section since July 1998. Prior to that she was employed as a Systems Librarian at Delta State University in Cleveland, Mississippi.)