

Digitizing the Administrative Records of the U.S. Department of Energy Office of Environmental Management (Em) and Office of Legacy Management (LM) Ohio Sites

J. Powell
U.S. Department of Energy Office of Legacy Management
11003 Hamilton-Cleves Highway
Cincinnati, Ohio 45243
USA

K. Williams, S. Walpole
S.M. Stoller Corporation
955 Mound Road
Miamisburg, Ohio 45342
USA

R. McKinney
Source One Management, Inc.
1225 Seventeenth Street, Suite 1500
Denver, Colorado 80202
USA

ABSTRACT

As former weapons sites close and are transitioned to the U.S. Department of Energy (DOE) Office of Legacy Management (LM), continued public involvement is essential for the successful turnover of long-term surveillance and maintenance (LTS&M) activities. During the environmental remediation process, public participation was a key factor in cleanup completion. The same level of commitment to encourage active public participation is true for the LTS&M activities at the LM sites, such as the Miamisburg Closure Project and the Fernald Closure Project. Community members participate in the transition and the decision-making processes for LTS&M as they did for the selection of response actions under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) cleanup process. [1] A key part of the post-closure activities for the Ohio Sites transitioning to LM from the DOE Office of Environmental Management (EM) is the preservation of site history and stakeholder involvement in the LTS&M activities that will continue during post-closure. In meeting the regulatory requirements of providing the CERCLA Administrative Record Reading Room for public access and to ensure that appropriate records are retrievable and available for all stakeholders, a decision was made to digitize the Miamisburg Closure Project and the Fernald Closure Project Administrative Records. This decision was, in part, based on the information and lessons learned from the digitization of the Rocky Flats Environmental Technology Site (RFETS) CERCLA Administrative Records (AR). The Ohio Sites effort was expanded to include the Living History Project from the Fernald Closure Project. In most cases, the CERCLA AR maintained by EM closure sites and transitioned to LM will provide adequate baselines for identifying and capturing the information required by LM for post-closure stewardship of the sites. The AR established under Section 113(k) [2] of CERCLA serves two primary purposes. First, the record contains those documents that form the basis for selection of a response action and comply with Section 113(j) [3]; judicial review of any issue concerning the adequacy of any response action is limited to the record. Second, Section 113(k) [2] requires that the AR act as a vehicle for public participation in selecting a response action. The AR is the body of documents that “forms the basis” for the selection of a particular response action at a site and contains historic information that has future study value by scholars, historians, regulators, and other stakeholders.

INTRODUCTION

The U.S. Department of Energy (DOE) Office of Legacy Management (LM) works to ensure that appropriate records are made available to stakeholders and the public in a timely manner and in accordance with appropriate security considerations. Scanning the various sites' Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) Administrative Record (AR) documents into a digitized format provides access from virtually any location while providing the security for disposition of the original documents to Federal Records Centers (FRC). Interested parties and stakeholders (local, regional, national, and international) who have an interest in the decisions made at the Ohio Sites can and will be able to electronically access the total CERCLA AR collections and, in some cases, supplemental information. Stakeholders and regulators are concerned that these sites are continually operated in accordance with published performance standards per agreed upon terms. These shared concerns can be quickly and efficiently addressed via the digitized CERCLA AR collection in conjunction with established Reading Room facilities. This arrangement provides open access to the data associated with potential migration pathways, clean up criteria, ongoing sampling, and the long-term surveillance and maintenance of the site.

Community awareness is vital to the continued protection of human health and the environment during LM's performance of LTS&M activities. Ensuring community awareness of the site and maintaining institutional controls will require outreach to new residents and future generations through effective dissemination of site information. Several LM sites have anticipated future land use (wildlife areas, industrial, etc.). The digitized collection can provide information on the assumptions regarding anticipated land use and institutional controls currently in place, based on the sites' proximity to urban, residential, commercial, industrial, agricultural, and recreational areas. The digitized CERCLA AR can include information on elements that are often under stakeholder consideration, cultural (e.g., historical sites, Native American religious sites), natural resources, environmental justice issues [3], location of on-site or nearby wetlands, geographic and geologic information, and location of wellhead protection areas, when developing assumptions about future land use. Stakeholder interaction is vital and allowing them to independently assess and make judgments about the significance of the environmental risks can be facilitated by the availability of the digitized site documents via the site-specific websites. In addressing the requirements of "all appropriate inquiry" rules, providing digitized, readily available information to stakeholders will assist in the distribution of data to support 40 CFR part 312, *Innocent Landowners, Standards for Conducting all Appropriate Inquiries*. [4] This openness and availability of information will help LM succeed and build credibility with all stakeholders. Throughout this process, consideration was given to ways to improve how information is communicated. Interaction with stakeholders begins the process of decision making that is meaningful to the various levels of stakeholders. However, stakeholder involvement is not limited to community leaders only. Community members who may be directly affected will be neighborhood businesses and homeowners. It has always been important to address key stakeholders early and make an effort to identify marginalized groups as well. Bringing together the right mix of stakeholders provides a comprehensive focus on the issues and informational needs of the stakeholders. This information is key to obtaining the full participation of stakeholders and in avoiding potential conflicts as well as identifying environmental justice [3] concerns, if any.

Leveraging both the tangible and intangible benefits of a cost-effective imaging system for digitization achieved a strategic gain for the Ohio Sites transition activities through enhanced decision-making, improved customer service (stakeholders), documented chain of custody, and heightened safety awareness using site historical knowledge. In addition, digitizing records reduced on-site paper volume by allowing paper records to be archived to an FRC. Reduction in the volume of the paper records enabled site organizations to minimize the square footage requirements for maintaining the CERCLA AR and the Public Reading room.

Digitizing records includes the following benefits:

- Reduces the space required to maintain paper records.
- Enhances search and retrieval capability for the stakeholders, LM staff, etc.
- Allows remote access by multiple users at the same time.
- Enables automated indexing and searching of key fields.
- Provides for full-text search capability when documents are digitized using optical character recognition technology.
- Allows rapid retrieval and increased availability of records for stakeholders and audits.
- Enables electronic back-up and recovery of documents in the event of a disaster.
- Enhances quality control of public offerings.
- Enables electronic sorting, grouping, document management, and search and retrieval.
- Converts digitized images to standard file formats for use with industry standard database software, providing a user-friendly system for stakeholder and regulator use across a broad spectrum of individuals.
- Electronic documents and data hosting permits secure web-based global access for all stakeholders.

Lessons Learned

Rocky Flats Environmental Technology Site (RFETS) was the first site to digitize their CERCLA AR. It was this initial phase that addressed the identification and submission of documents that would be part of the CERCLA AR file record compilation. A process reengineering study was conducted in an effort to streamline the AR process. In that analysis, it was determined that digitizing the AR file and providing an electronic index would both reduce the cost of processing AR documents and provide stakeholders with an electronic method to search and retrieve the documents. Project goals were to assist stakeholders in the retrieval of approximately 7.08 cubic meters of post Record of Decision (ROD) AR documents (i.e., for document review and/or provision of hardcopies); provide microfiche and index updates to the public repositories until such time as the AR file was digitized; and, to provide an electronic AR file for the public reading rooms.[2, 4] This process description is further delineated in 1-F78-ER-ARP.001, CERCLA Administrative Record Program.

Lessons learned from the RFETS digitizing experience included:

- Proprietary software requiring download and use of a non-standard viewer is not preferred. Many users experienced system incompatibilities or could not download the non-standard viewer.
- The software's search capability was not intuitive for standard users, which significantly reduced the usability of the system.
- The proprietary software was not readily customizable. Stakeholder and user requests to modify the display screens could only be accomplished through the software vendor. This eliminated the ability to customize display screens and provided constraints to customizing searches.
- Digitized searchable images in TIFF format are not desirable. In the TIFF format, each page of an individual document was displayed as a single image with a large file size. TIFF formats did not provide for individual pages of a document to be bound together into a single document. Large file sizes required long download times (e.g., over an hour) and in many instances the software would timeout before the download was completed.

Using the RFETS lessons learned, a standardized digitizing process and a web-based display solution was developed for the Fernald and Miamisburg CERCLA ARs. The final digitizing solution for the Ohio Sites CERCLA ARs included a centralized document imaging system and imaging center to avoid duplicative

and decentralized investments in imaging hardware and software in multiple (and likely incompatible) configurations.

ASSESSMENT

The Fernald Closure Project's 8.07 cubic meters and the Miamisburg Closure Project's 6.65 cubic meters of CERCLA AR documents were scanned into a standard file format (PDF Searchable Image) at a resolution of 300 dots per inch (DPI) and converted to Portable Document Format (PDF) for web-based display. PDF has become the de facto standard for display of documents on the World Wide Web and has been adopted by the National Archives and Records Administration (NARA) as an acceptable electronic archival format. Through the additional use of Optical Character Recognition (OCR) software, each PDF is made into a searchable image. This file type (formerly known as PDF Image+Text) preserves the look of the original scanned image while adding the ability for full text search. [6]

Before document preparation and imaging/digitizing began, a document assessment on the Ohio Sites CERCLA AR was performed by random selection and review of the original record copies using a checklist that included identifier (box or document number), location, description, type of media (paper, CD, disk, microfiche, other), and descriptions of paper record condition. Description of paper records included whether or not the documents were in folders, stapled, clipped, or otherwise bound, legible, sizes (216 x 279 mm, 216 x 356 mm, etc.), type of paper (copy, bond, onion, etc.), and correctly indexed.

Using the assessment to gather the baseline information for estimating and evaluating the project scope, a project plan was developed in conjunction with a schedule for completion of the digitizing projects from the Ohio Sites. Procedures were developed for the digitizing process that defined the removal order, identifiers, target sheets when oversize documents were involved, loose transmittal letters protocol, and the final organization within the records box. Each box was scanned and treated as a single batch. Preparation of each document included the removal of tabs, staples, binding pins, plastic bindings, and other fasteners. Pre-job identification of various document sizes and segregation of these documents was done to ensure appropriate staging for the oversized scanners and the continuous process flow throughout the digitizing room.

The table below describes each collection in detail (total pages, quantities of each size, and condition).

Table I. Estimated Records Collection Size

	Fernald CERCLA AR Collection	Mound CERCLA AR Collection
Total Page Count	710,991	586,971
% of Total	53.28%	43.99%
Pages 216 x 279 mm	688,211	569,875
Pages 279 x 432 mm	17,085	5,699
Pages > 279 x 432 mm	5,695	11,398
Pages 279 x 381 mm	0	0
Pages 254 x 279 mm	0	0
Pages 279 x 356 mm	0	0
Average Pages Per File 216 x 279 mm	56	56
Average Pages Per File 279 x 432 mm	54	22

	Fernald CERCLA AR Collection	Mound CERCLA AR Collection
Average Pages Per File > 279 x 432 mm	4	5
Pages Per File to be Scanned:	All	All
Paper File Status:		
File Condition:	Good	Fair to Good
Paper Size	216 x 279 mm 279 x 432 mm 864 x 1118 mm	216 x 279 mm 279 x 432 mm 864 x 1118 mm
Files Sorted:	By Document Number	By Document Number
% with Paper-Clips	0%	75%
% with Staples:	0%	90%
% with Binder Clips:	0%	10%
Double Sided Pages:	0%	TBD

THE TECHNOLOGY

- Paper files (up to 279 x 432 mm) are scanned at 300 DPI on a high speed Kodak i640 scanner.
- Large format drawings (up to 864 x 1118 mm) are scanned at 300 DPI using an HP Designjet 4200 scanner.
- Files are stored in web-optimized searchable image (formerly PDF Image+Text) Adobe PDF Format. Adobe Acrobat Reader is a free, universal software that lets you open, view, search, and print Adobe Portable Document Format (PDF) files. The PDF format allows anyone with a PC (23 platforms supported) to read files ranging in size from one page to several thousand pages.
- The database technology behind the CERCLA AR that allows users to search, sort and view the files is FileMaker Pro. FileMaker Pro has built-in web server support. The software can serve up to 100 concurrent web users, and has the ability to store one million tables, and 64 quadrillion records per table for a total of 6.4E+22 (64,000,000,000,000,000,000) records per database. Most single word searches in a title or subject field takes just a few seconds.
- FileMaker Pro Summary: Version 7 is a database program with significant graphical and networking capabilities. The most popular features are:
 - Equal functionality on Microsoft Windows (NT, 2000, or XP) and Apple Macintosh desktop computers, and support for simultaneous file sharing.
 - Self documenting graphical programming tools.
 - Excellent import/export, ODBC, and XML compatibility.
 - Network sharing with little or no additional programming on TCP/IP, IPX/SPX, networks, supporting up to 250 "logged in" users.
 - True client/server data processing, supporting Windows NT or 2000 Server, Mac OS X, and Linux data servers.
 - Self-contained web server for website data entry, presentation, or eCommerce.
 - Support for integration with Microsoft Internet Information Server web server.
 - Available with data tool for hand-held computers.
- FileMaker Pro Technical Specifications:
 - Hybrid Two-Tier Client Server Structure - Processes data in a two-tier method. The central server manages and indexes data, record locking, and monitors a user's activity. Client computers perform most of the calculations, data manipulation, filtering, display, and printing functions. FileMaker Pro maintains program code in a one-tier method. No program code or

templates are distributed to client computers. The central server maintains one central copy of the database code, and uploads the code to client computers as required.

- Data File Format - Uses a proprietary data format, and the data files are not compatible with other reporting tools (except using ODBC).
- Tables - Supports a maximum of one million tables.
- Data Table Size - Has a maximum data table size of 8 terabytes. Each table may store up to 64 quadrillion records.
- Security - Full account/password security with shared group level access levels.
- Programming Language - Uses a proprietary, hybrid graphical programming environment. There is no text editor or import/export capability for program code.
- ODBC - May function as both an ODBC client and an ODBC data source.
- ActiveX - Specific features can be controlled or accessed using ActiveX.
- HTML Serving - Has a built in http server for integrated database websites.

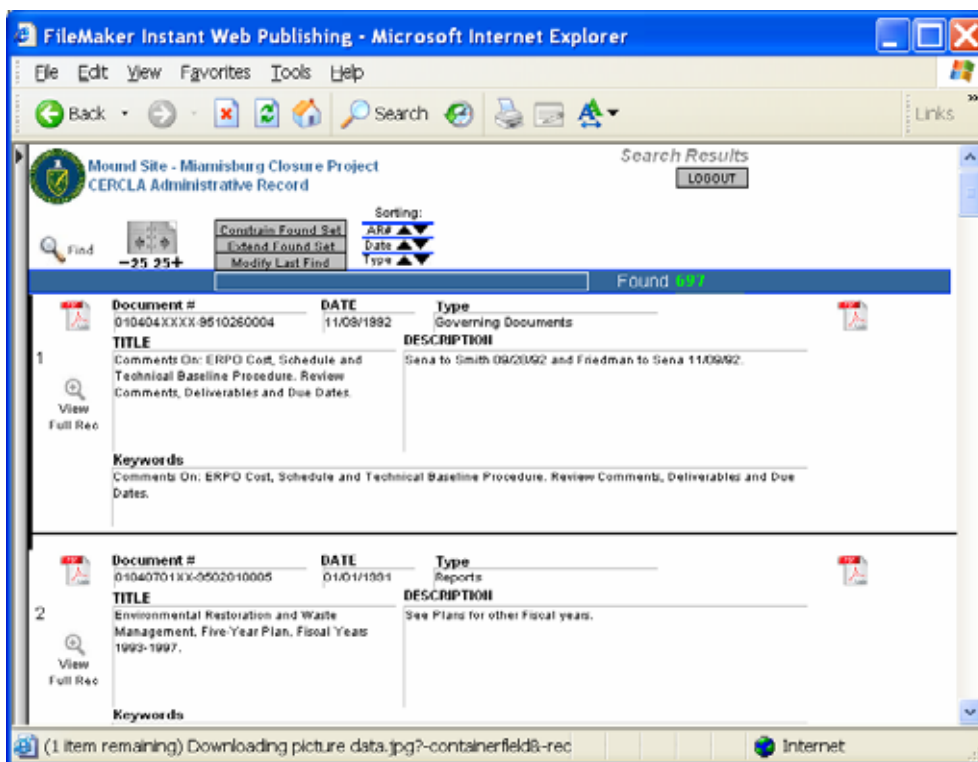


Fig. 1. Web Site Search Results Example

- Using FileMaker Pro technology, the LM CERCLA AR database was designed and output to a draft website in less than eight hours.

EQUIPMENT

- 1 ea - Kodak i640 Scanner, 100ppm – (2004 Association for Information and Image Management (AIIM) Hardware Mid-Range Best of Show Award) – Scans standard 216 x 279 mm up to 279 x 432 mm (max 304 x 660 mm)
- 1 ea - Kodak Care Kit - Extended service agreement - 1 year Mfg.#1073550
- 3 ea - Kodak Scanner Consumable Kit Mfg. Part# 108-4755
- 2 ea - HP Designjet 4200 Scanner – Large drawing scanner scans up to 1092 mm wide

- 2 ea - Adobe Capture 3.0 Cluster Edition (1 license per CPU) – Scan and convert documents to Searchable PDF format
- 1 ea - Kofax VRS Software for Firewire Production Scanning MFG#VP-P004-00U1 – Auto image enhancement software
- 3 ea - Dell Dimension 5100C Desktop PC – 3.2GHz, 2GB RAM, 160GB HD – 1ea PCs connected to the production Kodak i640 scanner, 2ea PCs processing images from the HP 4200 drawing scanners and also used for indexing & QC
- 1 ea – Standard barcode reader/scanner connected to PC for scanning document barcodes and ease of data entry
- DOE Rocky Flats web-based CERCLA AR Indexing System for presentation of CERCLA AR

THE PROCESS

The scanning process utilized support tools such as barcode readers (excess equipment from Rocky Flats) to minimize data entry upfront and to improve overall efficiency. [7]

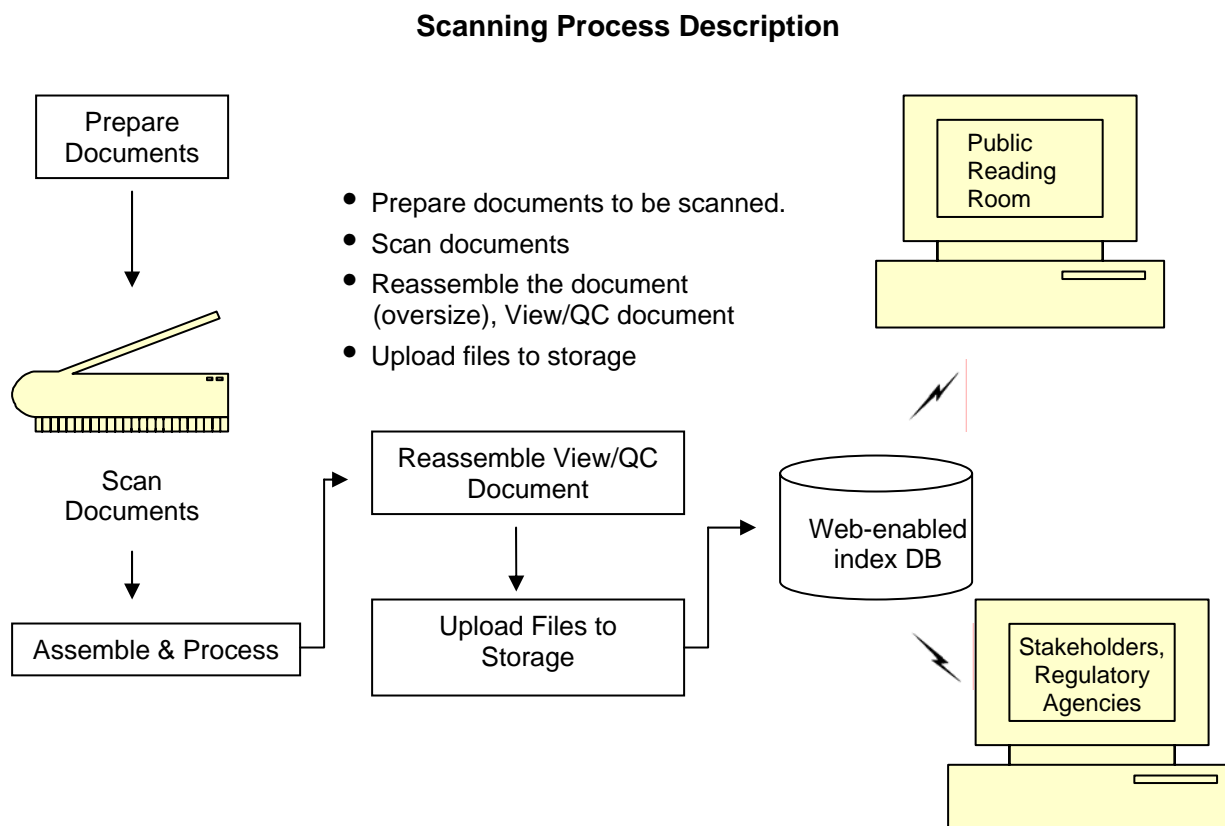


Fig. 2. Scanning process overview

Equipment and software was centralized and set up as a shared resource between the Ohio Sites, utilizing LM Records personnel from Miamisburg and Fernald. Equipment was purchased and installed in the building where technology infrastructure services such as telephone, computer security, networking, network storage and backup to store the scanned documents, already existed.

One workstation was designated for scanning documents and drawings, the second and third workstations were used for quality control of scanned image files and to verify, save, and store the searchable PDF files.

Document Preparation

Prior to digitizing, the CERCLA AR documents were prepared for the scanning process. Document preparation included the retrieval of boxes from on-site locations and, in some cases, the FRC. Each box held numbered folders containing the CERCLA AR documents. Each box was examined and the contents were verified. Any identified discrepancies were noted on a Document Pull Report, initialed and dated. Document and folder numbers were confirmed, again noting discrepancies, and appropriate contact for resolution was made. Once all issues were resolved, document preparation commenced and included:

- Removal of tabs, staples, binding pins, plastic bindings, and other fasteners.
- Examination of the document for oversize pages. Oversize pages up to 279 x 432 mm are turned on their side for ease of scanning.
- Pages that are larger than 279 x 356 mm are:
 - Removed in the order that they occur in the document
 - Clipped together with a sheet noting the document number
 - Page numbers are noted on folded drawings in the lower right-hand corner
 - Placed at the end of the document
 - A target sheet stating “Oversize Pages Located At End Of Document” is inserted at the beginning of the document.
- Examine the document for loose transmittal letters. If present, place letter in front of the document.
- Each document, after it is prepared for scanning, is placed in a new FRC box which has been assigned a box number by Records Management. Record the box number on the box as directed by Records Management and note the new box number on the Document Pull Report.
- Each box of documents scanned was treated as a single “batch” (one box = one batch.)

Scanning

Documents were scanned in batches and fed into the scanner so that the leading edges were aligned and centered under the feed module. This permitted the feeder to introduce documents into the scanner one at a time. Documents with missing corners, perforated edges, hole punches, irregular and curled edges, or damaged, were for the most part put through the scanner as is. In instances where the extent of damage was questionable, the documents were put into a clear protective sleeve and manually fed through. Optimal scanner performance is achieved using the recommended document specifications listed below. Scanning documents outside of these specifications may lead to undesirable results in terms of scanner reliability, image quality, and/or consumable life.

- Materials: Virgin, recycled, photographic papers, and clear protective sleeves meeting the appropriate size and thickness requirements.
- Paper Types: Bond, laser, inkjet, offset.
- Paper Weights: The elevator tray handles a broad range of paper weights from 45 to 200 g/m². The *Kodak Feeder Kit for Ultra Lightweight Paper* can handle paper weights from 25 to 75 g/m².
- Minimum Document Size: 6.4 x 6.4 cm. Documents as small as 5 cm (i.e., business cards) can be center-fed, in portrait orientation, one at a time.
- Paper Inks: All inks on the paper must be dry before scanning is started. This includes: Standard offset printing, inkjet printer, thermal transfer, and handwriting inks.

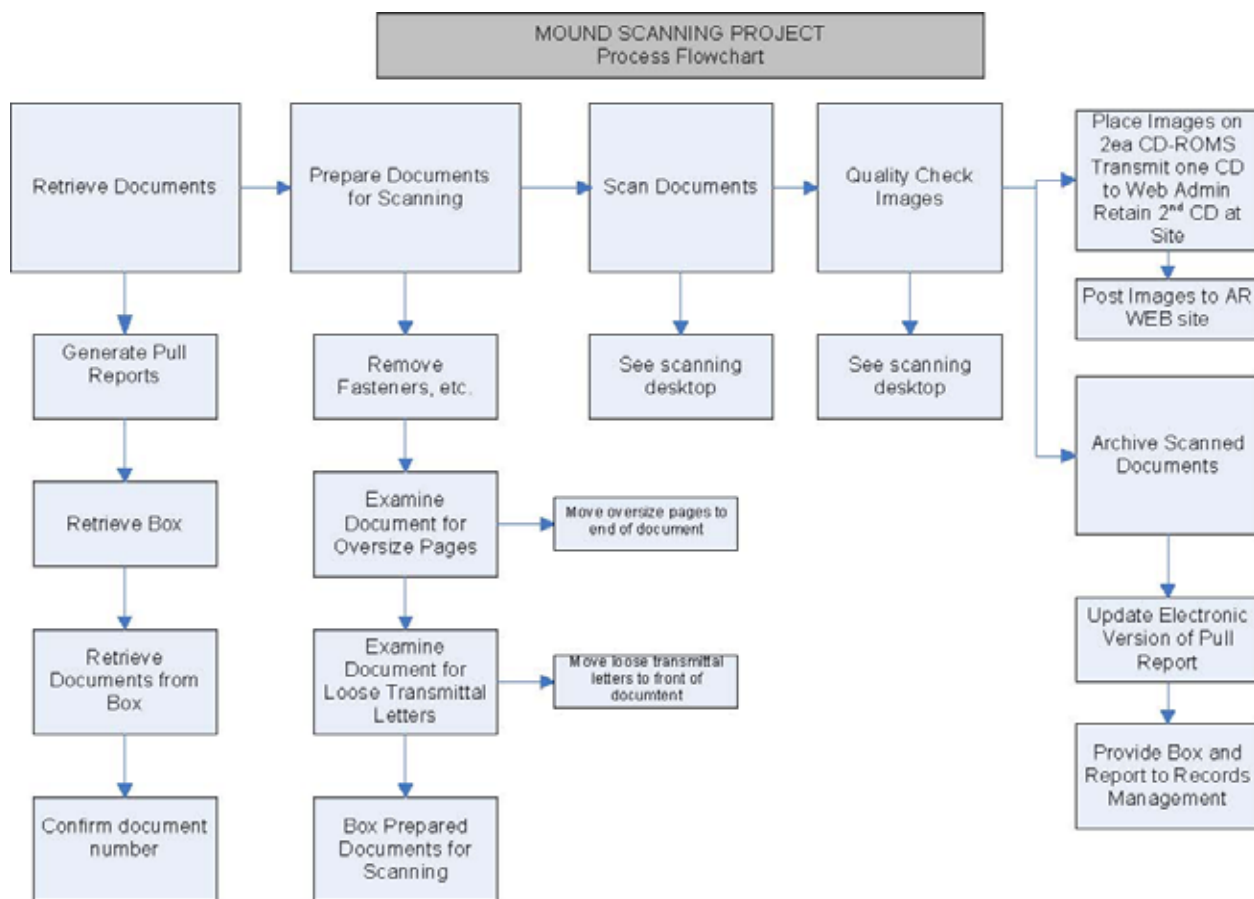


Fig. 3. Scanning process flow.

Assemble, Process, and Quality Control

Once scanning of a batch was completed, the original documents were reassembled and returned to the records box. The box was sent to the Quality Control (QC) workstation to begin the verification process where the corresponding PDF files were opened and scrolled through one page at a time. Visual confirmation of legibility and accuracy of the documents was performed. If one or more documents failed the QC check, the entire batch was returned to the scanning station, identifying the documents or pages to be rescanned. If the batch passed the QC check, the QC checkbox on the barcode sheet was marked to indicate that QC was completed. The document is then saved.

Upload Files, Archive, and Web Application

Moving the files from the Scanning folder on the server to the Archive folder on the server is the first step to creating the two CD-ROMs prior to uploading into the web application. To accomplish this, the batch name was located on the server, the folder name was selected and moved to the Archive folder. Once a batch is moved to the Archive folder, the box containing the paper records is transferred from the scanning area to the records storage area. At the completion of a batch, two CD-ROMs are created, one for the scanning operations area and one for the designated web master for uploading into the web application.

DIGITIZED ADMINISTRATIVE RECORD IMAGE WEBSITES

The CERCLA AR searchable PDF images are indexed and prepared for presentation to the public using FileMaker Pro software. Both Rocky Flats and the Miamisburg sites have been digitized and are available at the following web locations:

- Rocky Flats
<http://12.17.223.12/index.htm>
- Miamisburg Mound
<http://12.17.223.12/mound/index.htm>

The Fernald CERCLA AR is still in the digitizing phase of the scanning process. The searchable PDF files are expected online in a web-based environment in the spring of 2007.

REFERENCES

1. 40 CFR 300.415 (Removal Actions), 300.430 (Remedial Investigation/Feasibility Study and Remedy Selection), and 300.800-300.825 (Subpart I - Administrative Record for Selection of Response Action).
2. 42 *United States Code* (U.S.C.) 9613 [CERCLA subsections 113 (k) and (j)] and 40 *Code of Federal Regulations*, 300.415(n)(3)(iii) and 300.430(c)(iii), 01/23/2000
3. *Executive Order 12898, Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations*, 02/11/1994
4. 40 *Code of Federal Regulations*, Chapter 1, part 312, *Innocent Landowners, Standards for Conducting all Appropriate Inquiries*, 07/01/2006
5. *Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), as amended by the Superfund Amendments and Reauthorization Act of 1986 (SARA)*, 42 U.S.C. 9601, et. seq.
6. *Scanning Assessment*, Source One Management, Inc., 09/15/2005
7. *Mound/Fernald CERCLA AR Scanning Desktop Instruction*, Source One Management, Inc., 03/16/2006