Records Management in the Formerly Used Sites Remedial Action Program (FUSRAP)

G.N. Morekas, M.B. Pape
Dynamac Corporation
2275 Research Boulevard, Suite 300, Rockville, MD 20850
U.S.A.

ABSTRACT

The U.S. Army Corps of Engineers' (USACE's) performance of site investigation and remediation under the Formerly Used Sites Remedial Action Program (FUSRAP) requires the use of a records management system in order to effectively capture and manage data, document the decision making process, and allow communication of project information to regulators, congress, and the public. The USACE faces many challenges in managing the vast amount of data, correspondence, and reports generated under this program, including: management of data and reports in a variety of paper, electronic, and microfilm formats; incorporation of records generated by the Department of Energy (DOE) prior to 1997; ensuring smooth flow of information among numerous internal Project Managers and regulators; and facilitating public access to information through the development of CERCLA Administrative Records and response to Freedom of Information Act (FOIA) requests.

In 2004-2005, the USACE Buffalo District contracted with Dynamac Corporation to adapt the records management system developed for the Formerly Used Defense Sites (FUDS) Program to the records for the Luckey and Painesville FUSRAP sites. The system, known as the FUDS Information Improvement Plan (FIIP), was jointly developed by the USACE Hazardous, Toxic, and Radioactive Waste Center of Expertise (HTRW-CX), USACE Rock Island District, and several FUDS contractors (including Dynamac Corporation) in 2003. The primary components of the FIIP which address the challenges faced by the FUSRAP Program include: the development of a standardized document organization system; the standardization of electronic conversion processes; the standardization of file naming conventions; and the development of an automated data capture system to speed the process and reduce errors in indexing.

The document organization system allows for the assignment of each individual document to one of approximately 150 categories. The categories are based upon a combination of the USACE site approval process and the CERCLA process. Documents are assigned to categories based on their relevance to the approval of the site, historical or reference content, and generation during a phase of the investigation and remediation process. Within each category relevant to a CERCLA phase, subcategories exist to allow for the organization of correspondence, data, workplans, contracting information, technical reports, decision documents, and other associated documentation.

The electronic conversion system was developed by USACE to standardize the types of electronic files maintained on the USACE Project Information Retrieval System (PIRS). Prior to standardization, PIRS had been populated with hundreds of thousands of electronic files of various incompatible types, making systematic search and retrieval impossible. By dictating the

use of PDF files with standardized specifications, USACE ensured that all files were in the same format, and that the format is universally searchable and retrievable by probable users of the system. Similarly, the implementation of standardized file names, based on a combination of the site name, organization category, and a unique sequence number, facilitates linking of the individual files to indexing and search functions.

To automate the entire process, Dynamac developed a Document Management Database, which was adapted to the FUSRAP Program for the Luckey and Painesville sites. Prior to the use of a database, capture of document titles, authors, dates, organization category, and other information was done manually, taking substantial time and leading to numerous errors. The database, developed in Microsoft Access format, facilitates the speed of the data entry and reduces errors by building drop-down menus during the data entry process. For instance, once an author and their affiliation are entered once, the name and affiliation can be retrieved by entering the beginning of the character string, or choosing the name from the drop-down list. For projects involving the management of tens of thousands of documents, these functions can reduce the time needed for indexing by an order of magnitude. The database also speeds the process and reduces errors in several output functions of the system. Once document information is entered, the database is able to automatically generate document file names, document labels, a document index, and document cover pages in seconds. Without the use of an automated system, each of these must be developed manually, taking days, with the corresponding possibility of errors. The automation of the system not only eliminates errors, but ensures 100% correspondence of information between the PDF files, the index, the labels, and the cover pages.

A substantial advantage of the database, which contributed to its selection as the standard for the FUSRAP Program, was its development within Microsoft Access. Numerous off-the-shelf records management systems exist, with varying levels of functionality and ease of use. However, each of these requires separate purchase and training for the user. The choice of a Microsoft-based system for the FUSRAP Program allows the system to be operated without the purchase of additional software, and without substantial additional training for personnel.

INTRODUCTION

Dynamac Corporation (Dynamac) has participated in the development and implementation of a records management system to assist the U.S. Army Corps of Engineers (USACE) in the creation and management of Administrative and Permanent Records for their Formerly Used Defense Sites (FUDS) and Formerly Used Sites Remedial Action Program (FUSRAP) programs. In 2003-04, Dynamac was contracted by the USACE Hazardous, Toxic, and Radioactive Waste (HTRW) Center of Expertise (CX) to manage a portion of the Administrative/Permanent Records Initiative project (AR/PR project) for the FUDS Program. The purpose of the AR/PR project was to organize technical documents related to selected investigation and remediation projects, establish Permanent and Administrative Record Files for the projects, and make the newly established Records available in standardized electronic format. In early 2004, Dynamac managed several individual assignments to expand, modify, and automate the records management system as part of the FUDS Information Improvement Plan (FIIP), and to make it applicable to the FUSRAP Program.

The system has three primary components, including:

- Document organization, categorization, and naming;
- Capture of document information for indexing and searchability; and
- Development and organization of electronic files to facilitate access and searchability.

DOCUMENT ORGANIZATION AND NAMING

Two important components of any records management system are the organization of documents into a logical structure, and the development of unique document identification numbers to allow tracking and access of hard copy documents and electronic files.

In the development of the AR/PR project, USACE developed a hierarchical, CERCLA-based document organization system which provided numbered categories for documents. The system was included as Appendix K of USACE's EP 1110-3-8, *Public Participation in the Defense Environmental Restoration Program (DERP) for Formerly Used Defense Sites (FUDS).* This system divides documents into separate major categories based on the CERCLA phase with which they are associated, and into subcategories based on the document type. For example, Preliminary Assessment/Site Inspection (PA/SI) phase documents are filed within Section 1, Removal Response documents in Section 2, Remedial Investigation (RI) documents in Section 3, etc. Within each phase, subcategories exist for correspondence (for instance, RI correspondence is placed into category 3.1), data (category 3.2), workplans (category 3.4), reports (category 3.10), and a number of other subcategories. In addition to the CERCLA phases, major categories also exist for Community Relations documents (Section 8), congressional correspondence (Section 9), Freedom of Information Act (FOIA) documents (Section 10), and real estate documents (Section 11). In all, there are approximately 150 subcategories which can be used to organize the project documents.

In the development of unique document identification numbers, the decision was made to base the number on a combination of three elements: the project name or number, the CERCLA-based organization category and subcategory numbers, and a sequence number based on the chronological dates of the documents within the subcategory. This results in document identification numbers of the following format:

Painesville_03.10_0007

This name represents a document within the Painesville FUSRAP site file, which is categorized as an RI Report, and which is the seventh document, chronologically, in this section for Painesville.

The unique document identification number serves several purposes. It is listed on a label attached to the front page of every individual document within the official project file, ensuring that the number is clearly associated with the document. The number is also listed on the AR or PR index for the project, allowing cross-referencing between the individual documents and the

index. Finally, the number acts as the name of the electronic PDF file for the electronic version of the AR or PR file.

CAPTURE OF DOCUMENT INFORMATION FOR INDEXING

As part of the AR/PR project in 2003-04, Dynamac was required to create a document index in Microsoft Excel format for each project for which an Administrative or Permanent Record (AR/PR) File was developed. Under its own initiative, Dynamac chose to design a Microsoft Access database to capture the required information and automatically generate the required Excel index. Dynamac's internal database was modified frequently during the year to minimize database entry errors, increase efficiency, and provide enhanced output features.

Throughout the AR/PR project, many District office FUDS and FUSRAP personnel inquired about the ability to add new documents to the Administrative and Permanent Record Files developed by Dynamac. New documents could be assigned additional sequence numbers, labeled with manually generated labels, and added manually by USACE employees to the Microsoft Excel index provided by Dynamac. However, based on continued interest from the Districts in having a tool to develop or update Administrative Records internally, Dynamac received work assignments to improve the database, develop training and guidance materials, and provide the database to the Districts as a deliverable product. The new Document Management Database offers data entry and four automated output capabilities to allow for increased usability and efficient record keeping.

Database Purposes, Functions, and Capabilities

The purposes of the Document Management Database are to facilitate the collection of document titles, authors, dates, and other information required for the Administrative and Permanent Record Index and to give Corps' staff or contractors the ability to use the database in the continual building of the AR/PR Records for past, ongoing, and future projects. The three programs required for database operations are Microsoft Access 2000 (or more recent version), Microsoft Word, and Microsoft Excel.

The initial purpose of the internal Dynamac database was to facilitate the collection of document information required for the AR or PR index for the AR/PR project. The information required for this index includes the following items:

- Document title/description, author, recipient, and date;
- Document format (e.g., letter, technical report, e-mail, or other format);
- Project name and unique project number;
- CERCLA-based category assignment;
- Document's Chronological Sequence Number, and
- Administrative or Permanent Record Assignment.

Although originally intended only to capture required index information, modifications to the database permit the following capabilities:

- Maintenance of the information required for the Administrative and Permanent Record File Index;
- Automatic assignment of chronological sequence numbers to documents within their CERCLA-based categories;
- Automatic generation of document labels in Microsoft Word format;
- Automatic generation of AR or PR document indices in Microsoft Excel format; and
- Automatic generation of templates for project completeness checklists.

These capabilities allow the Districts' staff to add new documents or properties to existing or new Administrative and/or Permanent Record Files, maintain an updated inventory of all of the documents in the Record files, generate labels for new documents, and develop updated indices as new documents are added to the Record files.

Database Organization

The Database's functions are divided between four main Microsoft Access windows:

- Opening Main Menu Window;
- Enter Project Information Window;
- Document Check-In (data entry) Window, and
- Output Options Window.

The Opening Main Menu window primarily allows the user to view the list of previously entered projects and properties, create a project file in which records can be entered (Enter Project Information button), select from a list of previously entered District projects to open a particular project's database and add records (Input Document Information button), or to open the list of output functions that include sequencing, creating labels, creating indices, and creating checklists (Output Options button). Fig. 1., below, shows the Main Menu that appears when first entering the database for a District. In the example below, the Painesville database is shown.

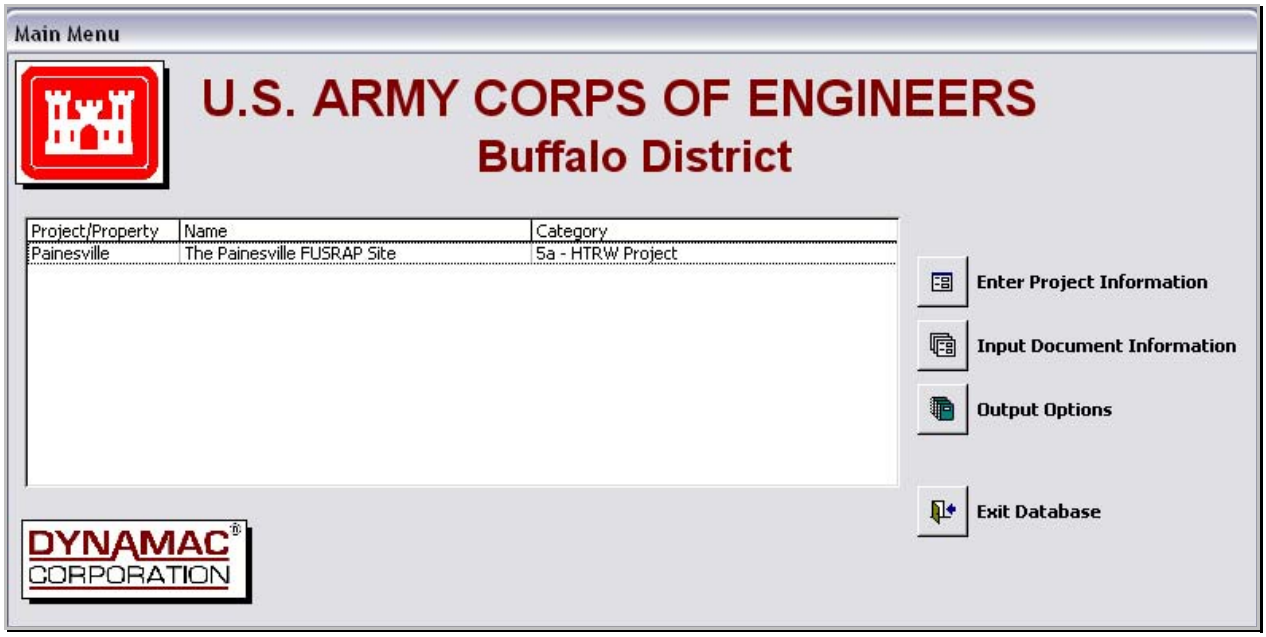

Fig. 1. Main Menu window

The user should select the Enter Project Information button when the user wishes to add a new or revise a previously entered project or property. The Project Information window simply captures the project name, project number, and project type, so that this information can be automatically accessed for the document labels, index, checklist, and other output features.

The Document Check-In window, shown in Fig. 2. has four primary areas:

- The top section, labeled "Project Information," contains the project or property name, project number (applicable only for FUDS projects), and category as entered on the Project Information window.
- The middle section, labeled "Document Description," is the location for entering the document-specific information, such as titles, authors, dates, CERCLA-based organization category, and other data.
- The section marked "Record" near the bottom allows the user to navigate through the individual project document records. Each document corresponds to a single record within the project. This field shows the total number of documents included in the system and the number of the record is being displayed.
- The "Return to Main Menu" button is located at the bottom of the Document Check-In window and will return the user to the initial Main Menu window.

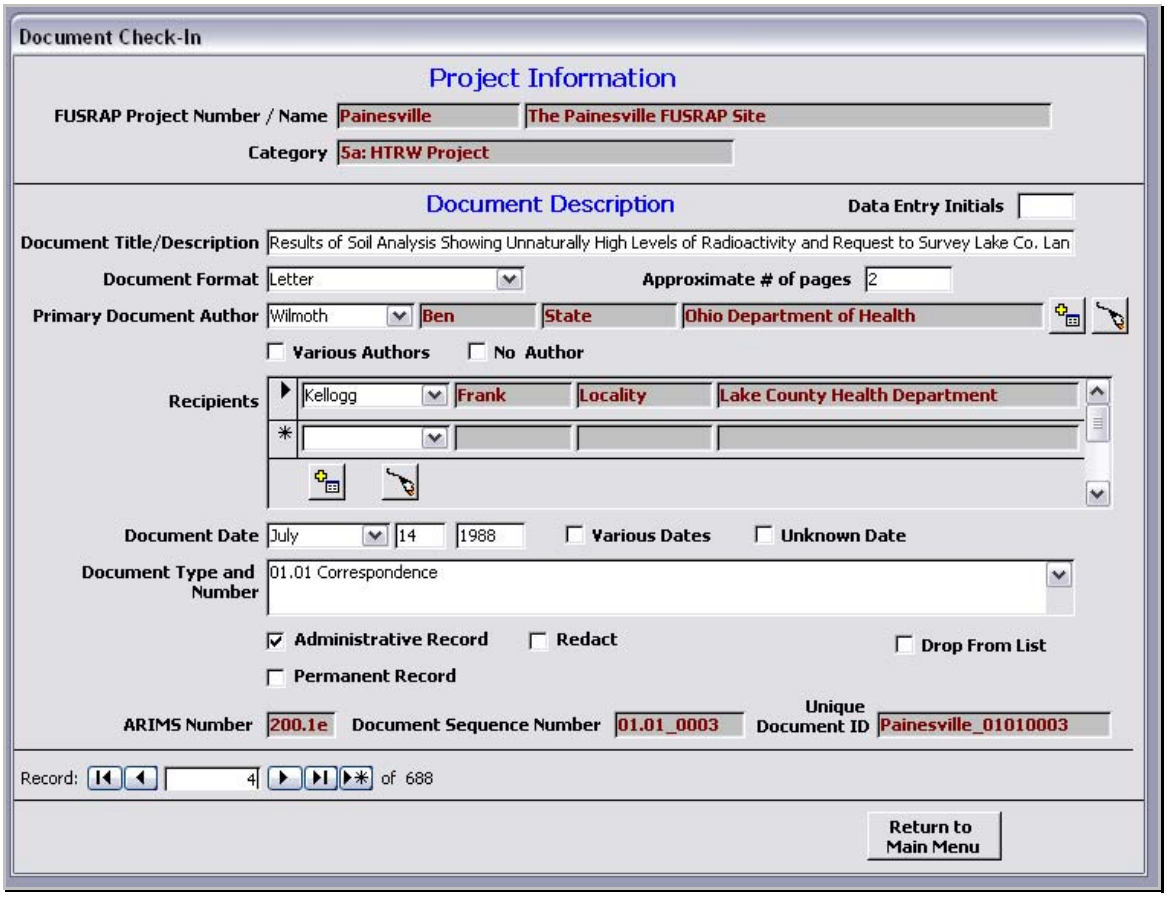

Fig. 2. Example Document Check-In window

One challenge driving the design of the database was the need to increase the speed and efficiency of the data entry process, which can be highly repetitive and error-prone. Dynamac designed the system to maximize the use of dropdown lists, which speed up the data entry by making it unnecessary for the user to enter the entire data string. Dropdown lists also reduce typographical errors by using pre-entered text with the correct spelling, rather than relying entirely on free-form entries. For data entry in fields with dropdown lists, the user can either use the mouse to display the dropdown list and click on a choice, or begin typing until the desired entry is automatically chosen. Some fields have dropdown lists that are built as the user adds names, such as the author and recipient fields. The user can add free-form entries into these fields. In the Document Management Database, once a name has been entered, the system automatically adds it to the dropdown list. The author and recipients dropdown lists are combined. Therefore, the addition of an author's name will appear in the dropdown list for recipients and vice versa.

The database also automatically sorts all documents in the selected project or property according to the CERCLA-based organization categories and then, within each category, chronologically from oldest to newest every time the user exits and reenters the project or property's Document Check-In window. Documents that are undated or that have various dates are sorted at the end of the CERCLA-based category to which the document was assigned.

Output Operations

The Document Management Database offers four output functions for each project or property. These output functions appear on the Output Options window, shown in Fig. 3., and include:

- Assigning Sequence Numbers;
- Creating Labels in Microsoft Word;
- Creating the AR and PR Microsoft Excel Indices;
- Generating project-completeness checklist templates; and
- Generating Administrative Record cover sheets

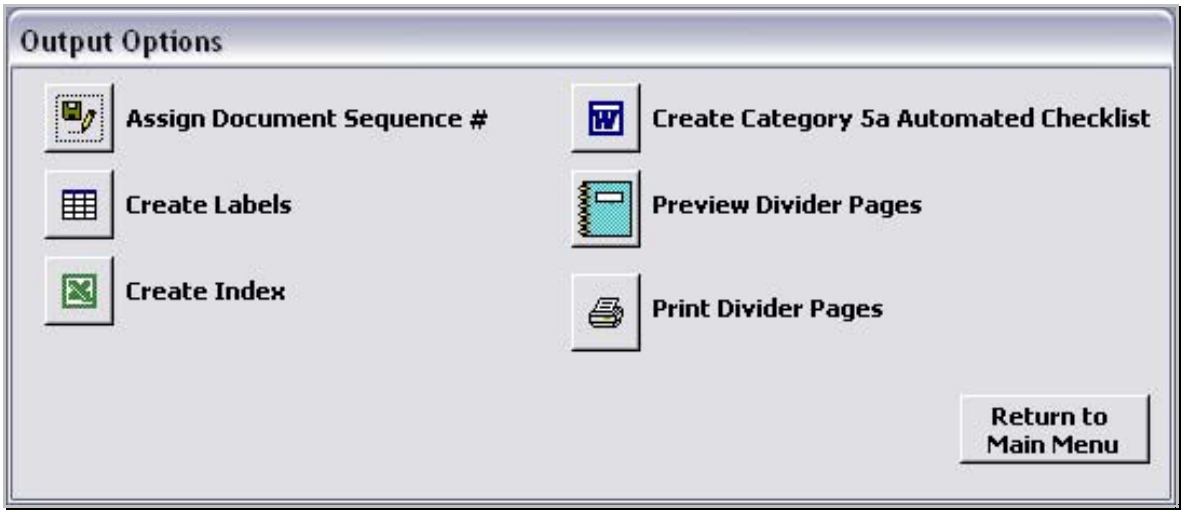

Fig. 3. Example Output Options window

Overall, the requirements for a unique sequence number, document label, record index, project-completeness checklist, and cover sheets for use in the hard-copy of the Administrative Record require the use of document specific data such as the title, author, dates, and CERCLA-based category. Prior to the inception of the database, each of these five requirements had to be filled manually by data entry into a Microsoft Word or Excel file. This resulted in the same information being entered separately into five separate files, taking an enormous amount of time, and presenting numerous opportunities for disparate information and typographical errors. The conception of the database was intended to require data entry of the required data elements one time, and then to access that information to complete the five output functions.

The "Assign Document Sequence #" button on the Output Options window automatically sorts all documents in the selected project according to the CERCLA-based organization category and then, within each category, chronologically from oldest to newest. The sequence number, along with the CERCLA-based category and the project name, serves as a unique identifying number with which to label and track each individual document. This feature eliminates the need to manually develop individual sequence numbers for each document for a project. Since the

sequence number is part of the label for each document, sequencing must be completed before labels can be created.

Prior to scanning, all documents receive a physical label applied to the front page. The components of the label are as follows:

- The ARIMS number, which is an internal Army tracking number required for permanent storage of documents. This number is generated automatically by the database, and is an indicator of whether the document is part of the Administrative Record (200-1e) or Permanent Record (200-1f).
- The Project Name-CERCLA-based Category-Document Sequence Number.
- Barcode, for automatic reading and input of document files names for electronic files.

Labels with the project name, the CERCLA-based K category, the sequence number, and the ARIMS number (AR/PR designation) are generated automatically in Microsoft Word format. Prior to the development of the database, the labels were generated manually in Word, which could take hours for a project consisting of several hundred documents. In addition, the manual process was highly error-prone, resulting in many mislabeled documents. By generating the labels automatically out of the database, the process can be completed within seconds, and the information on the labels is completely correct.

The "Create Index" button on the Output Options window automatically generates a separate Microsoft Excel document that is the index for the Administrative or Permanent Record for the project. Again, these Excel index was generated by manual data entry prior to the development of the database, taking a great deal of time and resulting in substantial typographical errors. By accessing the index information from within the database, the index can be generated within seconds, and the information is again completely correct.

Both the FUDS and FUSRAP programs require that checklists be completed for each project file, indicating the presence of certain documents in the file. The purpose of these checklists is to allow project managers and auditors to verify that all documentation required for project approval and to demonstrate compliance with CERCLA requirements is present within the project file. At this time, the database system does not automatically complete the checklist. However, it does generate templates for the checklist, in Microsoft Word format, and automatically places the project name and number information in the proper places on the checklist. The user must still evaluate the documents present and fill out the checklist.

Most of the USACE Districts maintain a hard-copy of the Administrative Record for each project in hard-copy format within their offices, and at a local library near the site. These records are usually maintained in multi-volume notebooks, with project-specific cover sheets on the cover of each notebook, and as the front-page of every individual document. These cover sheets, again, use the same title, author, date, and other information as the labels and index. Again, to eliminate the need for duplicative data entry, the database has been designed to automatically generate the cover sheets in Microsoft Word format.

DEVELOPMENT AND ORGANIZATION OF ELECTRONIC FILES

A primary motivation for USACE in implementing a records management system for their FUDS and FUSRAP programs was to identify means to store records to ensure their permanent integrity, yet continue to allow easy access to them when needed. To achieve this, USACE chose to standardize the types of electronic files, their files names, and their means of storage, searchability, and access. Prior to 2003, USACE had not directed their districts or contractors to maintain files in a standardized electronic format and naming convention, resulting in systems in which the documents existed, but could not be found or accessed. Starting with the AR/PR project in 2003, USACE directed that all future records be stored in PDF format, with very strictly defined scanning and properties standards, standardized files names with the unique document identification numbers, and Optical Character Recognition (OCR) features. These standards were chosen and established to facilitate document access by the ability to search for documents with a specified file name, and to provide text searching capability within each individual file.

As part of the development of the Document Management Database, Dynamac used the existing capabilities within Microsoft Excel and Adobe Acrobat to provide even greater access and searchability functions to the records. The system developed by USACE required an Excel index in which the user could read the document titles, dates, and authors, and then read the corresponding file name. The user would then sort through a hierarchical set of folders and associated PDF files, find the file of interest, and open it to read the document. Once the user had the PDF file open, they could then use the search function within Acrobat to identify specific text within the OCR text file. This system was usable, but required the user to sort through hundreds of folders and thousands of individual PDF files to identify the file of interest. In addition, the user could only search for text within a single document at a time.

Early in the process, Dynamac identified an easy, batch means to link the file name in the index directly to the PDF file. Instead of requiring the user to sort through the folders and files to go in and out of individual files, the user can now simply click on the link in the Excel index file, and the system immediately opens the correct PDF file. This allows the user to surf in and out of dozens of files within minutes. In addition, Dynamac accessed the cataloguing function of Acrobat to build catalog files for extremely large batches of PDF files. Using this function, a searchable catalogue of all text within a folder, on a CD, on a DVD, or on a network can be built to allow the user to search text through hundreds of PDF files at one time. The catalogue function of Adobe provides a list of links for all locations where the desired text string was found. Clicking on these links opens the appropriate PDF file, takes the screen to the page of interest, and highlights the desired text. Using this function, the users can move in and out of dozens of documents, looking for specific information, within minutes.

CONCLUSIONS

USACE's FUDS and FUSRAP programs made great strides in ensuring the future integrity and access to their records by establishing standardized electronic file requirements for the districts and contractors working on these programs. However, the new system was highly labor

intensive, required a great deal of repetitive and manual data entry, and still had limitations in the means of accessing and searching the documents.

Dynamac's initial impetus for developing the Document Management Database was simply to improve the speed of the development of labels, indexes, checklists, and cover sheets that were required to accompany the actual project files. By developing a system that could use a single set of document data to generate multiple outputs, Dynamac was able to eliminate all duplication in data entry, greatly speeding the process. In addition, by developing a data entry system that relied on pre-established and user-build drop-down lists, data entry was speeded even further. Both of these features also ensured that data within the labels, indexes, checklists, and cover pages was 100% consistent, and had a very low potential for typographical errors. Finally, the addition of the document access and searchability features for no additional cost improved the ability of the user to access and search documents by orders of magnitude.

Although these features were initially developed by Dynamac to facilitate internal operations only, they were adopted by USACE and added to the standard records management requirements in early 2004. The rationale of the USACE in adopting this system nationwide, instead of other off-the-shelf systems, was as follows:

- The added features could all be applied to records in batch, instead of file by file, allowing a large increase in accessibility with little extra effort;
- The systems use existing software, including Microsoft Office and Adobe Acrobat products that are already standard features on USACE computers; and
- The USACE personnel who will be required to use these systems in the future are already familiar with the functionality of the Microsoft and Adobe software, so development of the records and day-to-day user access to the records do not require any additional training.