ENVIROMENTAL
RESTORATION
PROGRAM

Electronic Document Management System
Analysis Report and System Plan
for the Environmental Restoration
Program

MANAGED BY
LOCKHEED MARTIN ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

UCN-17680 (8-85)
DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.
Electronic Document Management System
Analysis Report and System Plan
for the Environmental Restoration Program

C. Frappaolo

Date Issued—September 1995

Prepared by
Delphi Consulting Group
100 City Plaza
Boston, MA 02108-2106
under subcontract 28K-FGG28V

Prepared for the
U.S. Department of Energy
Office of Environmental Management
under budget and reporting code EW 20

LOCKHEED MARTIN ENERGY SYSTEMS, INC.
managing the
Environmental Management Activities at
Oak Ridge K-25 Site  Paducah Gaseous Diffusion Plant
Oak Ridge Y-12 Plant  Portsmouth Gaseous Diffusion Plant
Oak Ridge National Laboratory
under contract DE-AC05-84OR21400
for the
U.S. DEPARTMENT OF ENERGY

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
PREFACE

This *Electronic Document Management System Analysis Report and System Plan for the Environmental Restoration Program* was prepared in accordance with Lockheed Martin Energy Systems, Inc. systems development methodology requirements for feasibility study and requirements definition. This work was performed under Work Breakdown Structure 1.4.12.2.3.0.14.03 (Activity Data Sheet 8304, Information Resource Management). This document provides the Environmental Restoration Program with a detailed description of its current records and information management environment. Information provided in this document forms the basis for a recommended approach for automating key records and information management processes using electronic document management technologies. Components of a proposed solution, known obstacles, and a plan for executing an enterprise solution are presented.
CONTENTS

PREFACE ................................................................................................. iii
ACRONYMS AND INITIALISMS .................................................................. vii
EXECUTIVE SUMMARY ........................................................................... ix
1. BEFORE READING THIS REPORT ...................................................... 1
2. DEFINITIONS ....................................................................................... 2
3. PROCESS USED FOR THE PREPARATION OF THIS REPORT .............. 4
4. SITE ANALYSIS .................................................................................... 6
   4.1 THE CRITICAL SUCCESS FACTORS (CSFs) ...................................... 6
   4.2 OVERALL ARCHITECTURE AND DISTRIBUTION ................................. 7
       4.2.1 Paducah and Portsmouth .......................................................... 8
       4.2.2 Indexing and Storage .................................................................. 9
       4.2.3 Controlled Documents ............................................................... 10
       4.2.4 The AR Component ................................................................. 11
   4.3 RETRIEVAL/DOCUMENT ACCESS .................................................. 11
       4.3.1 Authoring ............................................................................... 13
       4.3.2 The System Infrastructure ......................................................... 13
   4.4 OTHER SITE CONDITIONS AND CRITERIA ...................................... 13
       4.4.1 Cultural Issues ........................................................................ 13
       4.4.2 Document Related Issues ......................................................... 14
       4.4.3 The Management Control Information System (MCIS) Infrastructure . 14
5. RECOMMENDATIONS AND OBSERVATIONS ....................................... 15
   5.1 THE NEED FOR ESTABLISHED PROCEDURES ............................... 15
   5.2 DOCUMENT MANAGEMENT WITH INTEGRATED FULL-TEXT RETRIEVAL (FTR) ................................................................. 17
       5.2.1 Benefits Derived .................................................................... 18
   5.3 HYPERTEXT AND COMPOUND DOCUMENT MANAGEMENT .................. 20
   5.4 EXCLUSION OF IMAGING AND RETENTION OF PAPER ..................... 20
       5.4.1 Disaster Recovery .................................................................. 21
   5.5 APPLICABILITY OF MULTIMEDIA .................................................. 21
   5.6 APPLICABILITY OF WORKFLOW .................................................... 22
   5.7 A PHASED IMPLEMENTATION .................................................... 22
   5.8 PERCEIVED OBSTACLES ............................................................. 26
       5.8.1 The Conversion Issue ............................................................... 26
       5.8.2 The Cultural Issues ................................................................. 26
6. CONCLUSION ...................................................................................... 28
FIGURES

Fig. 1. Overview of current ER system .................................................. 7

Fig. 2. Overview of proposed ER system ................................................ 23
ACRONYMS AND INITIALISMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Administrative Record*</td>
</tr>
<tr>
<td>ARF</td>
<td>Administrative Record File</td>
</tr>
<tr>
<td>BPR</td>
<td>Business Process Redesign*</td>
</tr>
<tr>
<td>C&amp;TS</td>
<td>Computing and Telecommunications Services</td>
</tr>
<tr>
<td>CDMS</td>
<td>Controlled Document Management System</td>
</tr>
<tr>
<td>CERCLA</td>
<td>Comprehensive Environmental Response, Compensation, and Liability Act of 1980*</td>
</tr>
<tr>
<td>CSF</td>
<td>Critical Success Factor</td>
</tr>
<tr>
<td>DMC</td>
<td>Document Management Center</td>
</tr>
<tr>
<td>DOE</td>
<td>U.S. Department of Energy</td>
</tr>
<tr>
<td>EDMS</td>
<td>Electronic Document Management System</td>
</tr>
<tr>
<td>EICMS</td>
<td>Electronic Information Content Management System</td>
</tr>
<tr>
<td>ER</td>
<td>Environmental Restoration Program</td>
</tr>
<tr>
<td>ERIS</td>
<td>Environmental Restoration Information System</td>
</tr>
<tr>
<td>FTR</td>
<td>Full-Text Retrieval*</td>
</tr>
<tr>
<td>IA</td>
<td>Information Assistant</td>
</tr>
<tr>
<td>IMS</td>
<td>Information Management Services</td>
</tr>
<tr>
<td>IRM</td>
<td>Information Resources Management</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LMES</td>
<td>Lockheed Martin Energy Systems</td>
</tr>
<tr>
<td>LMUS</td>
<td>Lockheed Martin Utility Services</td>
</tr>
<tr>
<td>MCIS</td>
<td>Management Control Information System</td>
</tr>
<tr>
<td>MTI</td>
<td>Midwest Technical, Inc.</td>
</tr>
<tr>
<td>NSF</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>OCR</td>
<td>Optical Character Recognition*</td>
</tr>
<tr>
<td>OREIS</td>
<td>Oak Ridge Environmental Information System</td>
</tr>
<tr>
<td>ORNL</td>
<td>Oak Ridge National Laboratory</td>
</tr>
<tr>
<td>PORTS</td>
<td>Portsmouth Gaseous Diffusion Plant</td>
</tr>
<tr>
<td>PGDP</td>
<td>Paducah Gaseous Diffusion Plant</td>
</tr>
<tr>
<td>PWT</td>
<td>Pacific Western Technologies</td>
</tr>
<tr>
<td>SAIC</td>
<td>Science Applications International Corporation</td>
</tr>
<tr>
<td>SDM</td>
<td>Systems Development Methodology</td>
</tr>
<tr>
<td>SGML</td>
<td>Standard Generalized Markup Language*</td>
</tr>
<tr>
<td>TCP/IP</td>
<td>Transmission Control Protocol/Internet Protocol (Ethernet™)</td>
</tr>
<tr>
<td>UT</td>
<td>The University of Tennessee</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
</tbody>
</table>

* Further explanation appears in "Definitions"
EXECUTIVE SUMMARY

Lockheed Martin Energy Systems, Inc. (LMES) has established and maintains Document Management Centers (DMCs) to support Environmental Restoration Program (ER) activities undertaken at three Oak Ridge facilities: Oak Ridge National Laboratory, Oak Ridge K-25 Site, Oak Ridge Y-12 Plant; and two sister sites: Portsmouth Gaseous Diffusion Plant in Portsmouth, Ohio, and Paducah Gaseous Diffusion Plant in Paducah, Kentucky. The role of the DMCs is to receive, store, retrieve, and properly dispose of records. In an effort to make the DMCs run more efficiently and to more proactively manage the records’ life cycles from cradle to grave, ER has decided to investigate ways in which Electronic Document Management System (EDMS) technologies can be used to redefine the DMCs and their related processes. Specific goals of this study are tightening control over the ER documents, establishing and enforcing record creation and retention procedures, speeding up access to information, and increasing the accessibility of information. A working pilot of the solution is desired within the next six months.

Based on a series of interviews conducted with personnel from each of the DMCs, key management, and individuals representing related projects, it is recommended that ER utilize document management, full-text retrieval, and workflow technologies to improve and automate records management for the ER program. A phased approach to solution implementation is suggested starting with the deployment of an automated storage and retrieval system at Portsmouth. This should be followed with a roll out of the system to the other DMCs, the deployment of a workflow-enabled authoring system at Portsmouth, and a subsequent roll out of this authoring system to the other sites.

It must be pointed out that the implementation of this proposed solution will necessitate the handling of nontechnical as well as technical issues. Managing user expectations, cultural resistance to change, and introducing a new level of acceptance of the role of the DMC will be as critical to the success of this project as choosing the right technology products. Efforts to handle these nontechnical issues should begin immediately under the direction of a project sponsor.

The next steps in this process are a determination of the specific features and functionality required of this integrated EDMS solution and an examination of available EDMS products. These are the subject of the EDMS Product Recommendations Report.
1. BEFORE READING THIS REPORT

This report is a combination of two tasks: Task I—EDMS Analysis Report and Task II—Comprehensive System Plan. It provides an analysis of the issues involved in ER establishing a new system through the deployment of EDMS technologies. This report will be followed and complemented by a Task III report in which specific solution alternatives will be reviewed and compared.

This report consists of five sections:

Executive Summary: Provides a synopsis of the findings and recommendations

Definitions: Establishes the report's vocabulary

Process Used for the Preparation of this Report: Outlines the methods and materials used

Site Analysis: Provides detailed discussion of the major factors and conditions which this report addresses

Recommendations and Observations: Defines the components of the proposed solution, states known obstacles, and outlines the execution of the solution

In order to fully benefit from the report, it is recommended that the reader pay close attention to the “Recommendations and Observations” chapter and become familiar with the terms listed in the “Definitions” during a cursory review of the report prior to a thorough second reading. The recommendations made in this report are based on findings to date and do not include considerations specific to product performance.
2. DEFINITIONS

Administrative Record (AR): A body of documents that represents the official public record regarding a particular response action. The AR is a subset of the records generated in the execution of an Environmental Restoration Program.

Business Process Redesign (BPR): A process, typically governed by a series of methodologies, which breaks down a business process into its smallest tasks and reconstructs the process, eliminating unnecessary steps and streamlining wherever possible. Sometimes referred to as reengineering, the goal of BPR is to evaluate established business models and redefine them so as to attain greater efficiency.

Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA): A federal statute that provides the statutory authority for cleanup of hazardous substances that can endanger public health, welfare, or the environment.

Compound Documents: Documents comprised of several modules each created separately but linked together in the document. The document components can be any combination of text, image, graphics, video, or voice.

Compound Document Management System: A system which tracks compound documents at the component level. A compound document management system manages the links between document components as well as the document itself. The links supported include live links, static links and intelligent links.

Document Management System: A database system for storing, tracking, and retrieving documents online. Document management systems provide library functions such as access security, document check-in/check-out, revision control, and usage audit trails.

Electronic Document Management System (EDMS): An online system which provides access to and control over any electronic document type (i.e., text, image, video, voice, CAD drawing, etc.). The EDMS is typically an integrated system comprised of any combination of imaging, workflow, full-text retrieval, document management, and multimedia technologies.

Full-Text Retrieval (FTR): A software- or hardware-based process that retrieves textual documents based on words, phrases, or concepts contained in the documents.

Hypertext: A text retrieval search methodology based on the associative memory process. Hypertext organizes information into sections known as nodes. These nodes can be linked together via macros that are automatically built by the system, based on user input.

Imaging: The process of capturing, storing, and retrieving information online, regardless of its original format, using micrographics and/or optical disk technologies.

Intelligent Links: Similar to live links, intelligent links are pointers to a relative file and its native application, but also track the relationship between document components. A change in one component may cause the automatic change in another.
Live Links: Pointers to a relative file and its native application. Live links provide the latest revision of a file each time it is accessed through a compound document.

Optical Character Recognition (OCR): Software technology that processes text within a scanned image and converts it into ASCII or word processing format.

Record: Information created by or for the LMES preserved or appropriate for preservation as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities or because of the informational value of the data regardless of physical form or characteristics. Records comprise the majority of documents created and maintained by or for the ER Program.

Standard Generalized Markup Language (SGML): A metalanguage that describes the language used by a text processing package to format text. SGML preserves text format in a document when it is transferred between platforms and applications. SGML takes a component approach to document formatting, treating the document as a series of related and linked bodies of subtext (i.e., paragraphs, titles, etc.).

Static Links: Pointers to a physical file. Static links embed a file into a compound document. The revision of the file is locked at the time the link is created.

Stemming: A process which reduces all query terms to their common root words and extends these to all possible permutations of the root word using standard syntactical rules.

Wildcarding: The use of characters known as wildcards in a query term which signify that any group of characters can be used to complete the term.

Workflow: A business process. Typically defined or viewed as a series of discrete work steps or tasks orchestrated to cooperatively develop a business goal or product. (Compare to Workflow).

Workflow: A proactive tool set for the analysis, compression, and automation of information-based business processes. A combination of software and design methodologies intended to redefine a work flow, exploiting automated procedures and electronic documents.
3. PROCESS USED FOR THE PREPARATION OF THIS REPORT

A series of interviews was conducted at various LMES sites and affiliated offices in Oak Ridge, Tennessee on January 30, 1995 through February 2, 1995. Additionally, teleconference interviews were conducted with branch sites, Portsmouth and Paducah. Interviews were conducted to determine areas of potential improvement and the appropriateness of EDMS technologies for the ER program. The primary objective of the interviews was strictly fact finding. A formal Business Process Redesign (BPR) was not conducted. However, many instances warranting a process redesign or reengineering did surface and are reflected in this report. The following individuals were interviewed:

Karen Andrews, Records Management Specialist (SAIC)
Marilyn Ayers, Integrated Project Management Manager (IMS)
Harry Boston, ORNL ER Program Manager
Tracey Brindley, Paducah Site Data and Records Manager
Stacie Burnham, ER IRM Systems Manager (IMS)
Dave Carden, Quality Assurance Specialist (DOE)
Jane Carr, Information Assistant (IMS)
Karen Cox, Administrative Record Coordinator (PWT)
Kathleen Fischer, Computing Specialist (C&TS)
Rick Forman, Senior Analyst (SAIC)
David Herr, OREIS Program Manager (C&TS)
Chris Hilemon, Computing Analyst (C&TS)
Bob Holmes, Technical Integration Program Deputy Director/
   ER Information Management Steward
Les Hook, ER Data Quality Program Manager
J.P. Hutson, Information Assistant (IMS)
Bruce Kimmel, ER Program Deputy Director
Becky Lawson, ER Records Manager (IMS)
Rick Lewis, Computing Specialist (C&TS)
Martha Knowles, Records Manager (Jacobs Engineering)
Debbie Matteo, ER Information Resources Manager
Clint Maynard, ER MCIS Manager
Jim McClanahan, Computing Specialist (C&TS)
Ivette Montalvo, Quality Assurance Specialist (DOE)
Gwen Patterson, Information Assistant (IMS)
Joe Payne, Information Assistant (UT)
Missy Sherrod, K-25 Site Editor (MTI)
Len Stenzel, Portsmouth Quality and Compliance Division Manager
Thomas Tallant, Technical Reports Analyst (IMS)
Ed Vazquez, Y-12 Plant ER Site Data Manager
Mike Ward, Assistant Vice President/Director Systems Integration,
   Software and Information Technology Division (SAIC)
Ben Watts, K-25 Site ER Site Data Manager
Robin White, ER Technical Integration Program Manager
Terry Whitlock, Information Assistant (UT)
Mark Wiens, ER IRM LAN Administrator (UT)
Rob Williford, Senior Analyst (SAIC)
Pam Wood, ORNL Document Management Liaison (IMS)
The individual findings in each interview are not detailed in this report. Discussion of the overall findings, however, is presented in the “Site Analysis” and “Recommendations and Observations” chapters of this report. Analysis was executed using Delphi Consulting Group’s internal knowledge base and several BPR analysis tools developed by Delphi Consulting Group.
4. SITE ANALYSIS

LMES has established and maintains Document Management Centers (DMCs) to support the Department of Energy (DOE) related Environmental Restoration Program (ER) undertaken at ORNL, K-25 Site, Y-12 Plant, and two sister sites: Portsmouth, Ohio, and Paducah, Kentucky. The DMCs were created in response to CERCLA legislation which places requirements on the type of documentation that must be created, maintained, distributed, and managed over 75–100 years. The role of the DMCs is to receive, store, retrieve, and properly dispose of records. These records consist of correspondence, technical reports, data packages, photos, drawings, and various other documents.

In an effort to make the DMCs run more efficiently and to more proactively manage the records’ life cycles from cradle to grave, ER has decided to investigate ways in which EDMS technologies can be used to redefine the DMCs and their related processes.

4.1 THE CRITICAL SUCCESS FACTORS (CSFs)

The focus on the DMCs is directly linked to CERCLA. The need to adhere to this government mandate has helped to define the basic business requirements or critical success factors (CSFs) associated with this project. The identification of CSFs is imperative in designing an EDMS solution because they dictate the manner in which the EDMS solution must be deployed. In the case of ER, the CSFs are as follows:

- **Tighten control over the ER documents.** Despite CERCLA requirements, the ER record creation and retention practices have been relatively informal and lax. ER must tighten control over these documents, from point of creation to archival storage. To this end, the focus is on establishing a central control function and abating the satellite mentality that has developed over the last few years. Thus, rather than having individuals and branch sites responsible for their documents, the Central ER DMC will become solely responsible for the management of ER records.

- **Establish and enforce record creation and retention procedures and rules.** This CSF is an adjunct to the first. As part of tightening controls over the record keeping system, rules need to be established which clearly outline which documents need to be created, which need to be kept, by whom, and for how long.

- **Become more responsive to user requests for information.** Information must become readily and reliably available to a variety of individuals from technical advisors to DOE auditors. Each request should result in a conclusive finding if the document was ever created. Retrieval should be accomplished in minutes (i.e., 1–5 minutes per request).

- **Demonstrate improvement to the ER system in a timely fashion.** It is desirable to have a viable solution (for at least a piece of the system) within six months.
4.2 OVERALL ARCHITECTURE AND DISTRIBUTION

ER records are created by a variety of DOE, LMES, and other subcontractor personnel. Each document type has its own associated purpose and life cycle, but generally, the documents must be maintained and available for a period of 75–100 years.

In the current approach, there is a central DMC located physically at the K-25 Site. This DMC was established to act as the single "record copy" DMC. The establishment of this DMC, however, was preceded by four satellite DMCs which continue to operate and which remain the focus of this system for many users. The central DMC located at the K-25 Site is intended to function as a DMC for all Oak Ridge sites. There are two satellite DMCs at ORNL, one at the K-25 Site, and another at the Y-12 Plant. Within all five of the DMCs, there is a basic consistency in the approach and tools used to maintain the document collection. Fig. 1 provides an overview of the basic architecture used at the central DMC and each of the satellites. Where applicable, the differences between these sites are noted in this report.

Fig. 1. Overview of current ER system. This schematic provides an overview of the general approach to document creation, storage, and retrieval within the current ER system. Authors create documents in a procedure that involves a series of steps and subprocesses. Most notably, the document must go through a clearance subprocess before it is released. Based on policy and procedure as well as requests from the AR system, some documents are specifically created for and routed to the AR system. Upon release, some documents may be revision-tracked and distributed through the control subprocess. In any case, the documents are routed to the DMC (central or local) for long term storage and ongoing access. Access is based on user requests.

Each DMC utilizes an askSam database to track the individual documents in their respective collections. (See the "Indexing and Storage" overview in this chapter of the report for more detail.) The documents are maintained as paper files in each instance. The procedure used to retrieve the documents is also similar from site to site. Each DMC has its own Information Assistant (IA). Documents are submitted in one of three ways: staff use a standard distribution list,
subcontractors/authors put the DMC on their distribution lists for correspondence, or project managers submit documents at the end of the entire project. Despite the presence of these DMCs and the IAs, submission of documents is voluntary. In many cases, project managers are retaining documents in their local files. Getting individuals to surrender control of their documents to the DMCs is improving but is still an issue at each site.

Despite the intent of the central DMC to function as the sole “record copy” DMC, only 20% of the ER document collection is stored and maintained there. The remaining 80% is in the satellite DMCs and in nonofficial repositories (i.e., personal file cabinets). Documents continue to be added to the DMC collections at a 20% average increase per year. K-25 Site estimates that 100 documents per week are being added to their DMC. Retrieval requests average approximately ten per day at each site.

4.2.1 Paducah and Portsmouth

Additionally, the sister sites in Portsmouth and Paducah maintain their own DMCs. While there is a desire to bring these sites into compliance with future directions established at Oak Ridge, currently these sites operate independently in different manners. The general scope of the sites (i.e., types of documents created, retention schedules, types of authors, and needs for retrieval) are similar, but the methods of storage, indexing, and retrieval vary. Unless indicated in this report, the needs of these sister sites is assumed to be the same as the needs in Oak Ridge.

4.2.1.1 Paducah

Paducah operates similarly to the Oak Ridge sites. There is an established DMC with a records manager and support staff of four. Files are maintained in paper, but indexed and tracked online. Where the Oak Ridge sites use a barcode to track documents (see the “Indexing and Storage” overview for more detail), Paducah uses a computer generated number as its unique document identifier. Instead of an askSam-based system, Paducah uses a document management system, developed by third party contractor, SAIC, which is based on the ORACLE relational database. Despite these differences, the indexing approach is fairly similar to that used in Oak Ridge. There are 13,000 documents indexed thus far. There are approximately 15 requests a day for documents. Controlled documents are handled by the document authors. (See the discussion on “Controlled Documents” below for more information on this subcollection of documents.) Similar to operations at Oak Ridge, an individual is assigned to administer the AR. (See the discussion on “The AR Component” below for more information on this subcollection of documents.) Requests for documents are handled in person at the DMC, and documents must be viewed there. If the end user wants to view the document at his/her desk, a copy must be made of the original. Extensive copy jobs are sent to the copy center at Paducah. Somewhat complicating the records management situation in Paducah is the fact that ER records management is handled by two contractors, LMES and Lockheed Martin Utility Services (LMUS). LMUS handles its own set of records, in excess of 13,000 documents. These documents are slated to be brought into ER’s control and indexed in the Oracle-based system. The volume of these documents will exceed the current physical capacity of the document vault.

The bottom line regarding Paducah however, is that the situation is basically under control for the time being. Documents are retrieved, on average, within one hour. They have had several successful audits. A new document management system developed in Microsoft’s ACCESS will allow users to retrieve documents online from their desktops. Approximately 180 Paducah users will be put into this network.
4.2.1.2 Portsmouth

The situation at Portsmouth is radically different. There is no established DMC at Portsmouth. Policies and procedures are in place similar to those at the other DMC sites, but there has been little effort to enforce them. There is minimal indexing of documents at Portsmouth. Although askSam was brought in recently, less than .5% of the documents are indexed in this system. Project files are stored in boxes and file cabinets. The search process is largely manual.

Despite the presence of 2,000 cubic feet of documents and the lack of a proactive document management system, user satisfaction is relatively high. This is predominately due to a perceived high level of complacency. The major source of complaints is the lack of physical space for more boxes. Portsmouth is one site, and thus by default, the single DMC is a central resource. But the DMC is ineffective since there is a strong cultural reliance on the individual document collections maintained by project managers. There are 200 people involved in ER at Portsmouth. As a result of this approach to document management, Portsmouth has been subject to several unfavorable audits by the DOE. In response to these audit findings, Portsmouth has allocated funding and established a subcontract with SAIC to create a solution to this situation. Portsmouth plans to address many of the audit findings by the end of FY95.

4.2.2 Indexing and Storage

4.2.2.1 Reports, Correspondence, and General Documents

When documents are submitted to the DMCs, the IA checks the askSam database to see if the document was previously submitted. If not, they are barcoded (labeled) and entered into the askSam system. The barcode is assigned by the indexer in sequential order. The barcode is used from that day forward as the unique document identifier. If the document was already coded by a satellite DMC, the central DMC will override the number with a newly assigned one. There is a “standard” template for tracking the document in askSam. Approximately 25% of the template allows for site-specific customization, however. The index tracks the document via values such as date entered, project name, the document type, title, author, recipient, etc., in addition to the barcode number. To provide a level of content-based retrieval, keywords are used. There is little to no control or creativity, however, in the use of keywords. The keywords are used sporadically and not consistently from site to site. In most cases, the IA is responsible for keyword assignment. A thorough assessment of the document’s content is not made in order to derive the keywords. The keywords are determined after a glance at the document’s first page. Because the IA is not technically oriented, no subject matter expertise is imparted into the assignment of keywords. In some instances a short abstract or comments are also supplied. These are generally provided by the submitting author within the content of the document.

On average, it takes 1–5 minutes to index a simple document such as a letter, and 5–10 minutes to index a more complicated document such as a report. While a document is awaiting index processing, it is virtually inaccessible. This time element can be significant considering that a document is not necessarily indexed as it is received in the DMC. This queue time can add approximately another day to the index process.

Once indexed, the document is filed. A variety of filing schemes are used across the DMCs. At central, the documents are predominately filed by barcode in sequential order. But there are some exceptions to this rule. At the satellites, the most popular approach is to file by barcode within project folders. They are grouped by project so a project’s entire document collection can be easily accessed.
Some documents within the ER collection make reference to other documents. There is no proactive system in place for tracking and controlling these relationships. If one document references another, the user must manually initiate another search to locate and retrieve the referenced document. In the overall scope of indexing and tracking, however, this does not pose a major problem as cross-reference-based retrieval does not occur frequently.

It should be noted that several individuals stated a problem exists with misfiled and missing documents. No statistics are available which quantify this problem, but it is none-the-less being considered a symptom of the current system that must be addressed.

4.2.2.2 Photos and Videos

Photographs and videos are indexed at the project level only. These are stored in separate systems, utilizing boxes. Retrieval is limited to associated project name.

4.2.2.3 Data Packages

Data packages are a subset of these document collections. They contain the raw data used to come to the conclusion reached in the documents. These packages of information are received from laboratories in two formats: paper and diskette. The diskettes are loaded into another online tracking system, OREIS. OREIS was created by DOE mandate. It tracks all analytical data. Data packages are one component of this system. OREIS is an ORACLE SUN-based system.

Data packages are handled differently at each site. The Y-12 Plant uses the electronic version of the data package as the record copy. The K-25 Site gives the diskette to the appropriate subcontractor who submits the contents to the local DMC at the conclusion of the project. ORNL forwards the data packages in boxes to their respective DMC for tracking. Copies are made prior to this and supplied to the data validators. There is no consistency between the various sites as to how the data packages are handled. Some sites submit the diskettes to the DMC, others do not. There is a general lack of consensus regarding the importance of the data package and how closely it should be managed. There is equal disagreement regarding the need to link the DMC tracking of the data packages to the raw data from the package that is ported to the OREIS system. But, there is general consensus that the data packages are not accessed after the first six months of their existence. Although there is no conclusive evidence that an auditor would ever need or want to trace an OREIS record to its data package in the DMCs, there is a legal requirement to maintain the data packages associated with OREIS data. The data packages are stored in their original boxes in the DMCs and are indexed by project, date, and location. Indexing is done at the box level, not by individual content.

4.2.3 Controlled Documents

Certain documents that are managed by the DMCs need to be controlled. A controlled document is one that requires revision control. Recipients of the controlled document are ensured that they will have the latest revision at any point in time. There is not a clear definition of what documents need to be controlled. In most instances, the author can decide if the document needs to be controlled. It should be noted that the system discussed here does not include the controlling of procedure manuals which are controlled in a separate system.

The distribution and tracking of controlled documents at the Oak Ridge sites is done in a system that is separate from the askSam database. The Controlled Document Management System (CDMS), was developed specifically to manage the control process. CDMS is operated by one individual, but two
backup people are being trained. CDMS has been available for four years. It is a database system with an integrated report writer. Approximately 2% of the total ER document collection is subject to control. Despite this low percentage, the control process is important because it is a government compliance issue. There are approximately 52 documents in the CDMS currently. Input is sporadic but increasing.

If an author requests that a document be controlled and forwards that document to the DMC, the IA uses CDMS to track who the recipients are, assign them a control number, and distribute the document with an automatically generated cover letter. These letters are to be signed by the recipient and returned as a record of receipt. Returned transmittal letters are retained forever to prove receipt. If the recipient doesn’t respond, they are sent three more letters and a final phone call. If they fail to respond to all notifications or specifically express no further interest in the document, their name is removed from the controlled document distribution list. The DMC is always placed on each controlled document’s distribution list.

When received at the DMC, the controlled document is stamped “controlled” and indexed in the askSam system, as any other document is. Thus, controlled documents are tracked in two systems, CDMS and askSam. Individuals who are not recipients of controlled documents but access them at the DMC may not take a copy of the document, unless they receive permission from the author and are added to the controlled document distribution list.

4.2.4 The AR Component

A subset of records known as the Administrative Record (AR) is managed by ER since it is the integrating contractor. These records are maintained as testament to a particular response action. The focus of this study is not concerned with the AR, as it is outside of this task’s domain and viewed as a separate albeit related document collection. The AR system must be surveyed, however, in order to appreciate the relationship between the DMCs and the AR system. Furthermore, a product being developed to manage the AR will be considered in the Task III report as a possible solution for the DMCs.

There are two sets of files maintained as part of the AR system: the Administrative Record File (ARF) and the Administrative Record (AR). The AR system currently utilizes an InMagic-based system to track the ARs. This InMagic system is used similarly to the manner in which askSam is applied to the ER records. A system known as ERIS is being created not only to track the ARs but also to make the document images available online through a combination of index values and full-text retrieval of document contents. Upon implementation of ERIS, the use of the InMagic-based system will be discontinued.

Integration between ERIS and the DMCs can be kept at arm’s length. The areas of overlap between the two systems (i.e., when do documents get forwarded to the AR and what format should they be in?) should be addressed in the design of the DMC solution. But other than this, they are separate systems.

4.3 RETRIEVAL/DOCUMENT ACCESS

ER documents are requested by a variety of individuals at each of the DMC sites. The level of activity at each site is not proactively monitored, but an average of ten retrieval requests per day per site was generally accepted by the IAs at each of the sites. It is anticipated that if this project is successful, the number of requests for documents would likely increase as users became acclimated to and gained greater faith in the DMC’s document management system. Based on a variety of factors, including the degree to which the askSam index can be used and the availability of local copies, retrieval can take anywhere from one minute to hours or days.
The requesters of the documents vary from site to site and in some instances by the related project. The requesting community is comprised of validators, analytical coordinators, project leaders, technical assistants, risk assessors, editors, authors, secretaries, engineers, subcontractors, ER management, and auditors. Requests can come in a variety of ways: e-mail, fax, and phone. Phone requests comprise the overwhelming majority of the requests.

All retrievals are executed by the IAs at the DMCs. Based on the requester’s input, the search is based on index values in the askSam database. Once the document(s) requested are identified in the askSam system, the IA must pull the appropriate paper file and make a copy. In some instances, the IA will maintain extra copies of “popular” documents in a copy folder, sorted by document identifier. In these cases, this folder is checked prior to accessing the “original” in order to save copy time. If the IA cannot locate the document(s) in the local DMC, phone calls are made to the other IAs to determine if the document is in the other DMCs. There is a plan to link the separate askSam indexes across the DMCs, but this has not yet happened.

The success rate in locating documents varied among those interviewed. Though no one claimed 100% success, some reported success rates as low as 50%. The overall attitude regarding the success of retrieval was most complacent. One individual commented that if he could not find the document(s) he was looking for, he labeled this as a “data gap” in his work and proceeded.

In spite of the askSam system, which has increased retrieval success over a manual approach, there are still instances of user frustration caused by not knowing how to succinctly state a request within the context of the askSam index. Typically, a request for a document yields several documents meeting the search criteria. In these cases, the IA prints out the askSam index records for the candidate documents and asks users if they can identify which ones they want to see. In some cases, the documents have to be pulled and shown to the user before this decision can be made. Instances where as many as 30 documents have been pulled and copied only to have 29 tossed away were reported. The need to see the documents is based primarily on the lack of detail available in the index records. For example, knowing that a document discusses mercury does not let the user know whether mercury is a main or peripheral topic within the report. In any regards, retrieval is most often an iterative process in which users submit a query and get back a set of candidate documents. A second query is executed on that set to further narrow the selection. This process repeats until pertinent documents are identified.

Despite the presence of the askSam index, retrieval is fairly dependent on user memory (i.e., I know how to pose the query because I remember the title of the document that I need to see). For this reason, newer users of the system experience greater levels of frustration with the system, since they lack the insight that experience provides. One user commented, “The current system is good if you know what you are looking for, but if you don’t always know exactly what you need, there is little help in finding it.”

Concern was expressed with the validity and usefulness of keywords. While nearly everyone is using them as a means to aid retrieval, reliance and faith on them is low. Because the assignment of keywords is superficial, the intelligence that can be derived from them is not high. Some users expressed concern that keywords are not fully indicative of document content and value, and that keywords can become dated. For example, one individual recalled a time when a request was made for all RI reports that discussed mercury. A list was produced, and from this, documents were selected. But the user was never sure that indeed all of the appropriate RI reports were identified. The user was certain that there were probably other reports that discussed mercury, but were not keyworded on this term because it was not a major theme of the document. In these cases, the user accepted the fact that the documents were simply “inaccessible.”
On a related note, IAs are using wildcards as a means of increasing the effectiveness of searches (i.e., expanding the search terms to all its possible permutations, such as plurals, verb form, etc.). Though this is a valid approach to increasing the effectiveness of the current retrieval system, in reality wildcarding is not a preferred method for achieving this functionality. (This matter will be discussed further, and alternatives presented in the Task III report.)

Lastly, no facilitated means exists for accessing documents that are still in process. Because a document is not indexed until it is finalized and submitted to the DMC, if a user needs to see such a document, the user or the IA makes a series of phone calls to probable authors in an attempt to locate the document. In this scenario, retrieval is seriously limited by the users knowing exactly what they are looking for. For example, there would be no practical way to locate documents in process that discuss mercury contamination of water, unless the user knew precisely that these documents were being worked on and by whom.

4.3.1 Authoring

The creation cycle of a document is currently viewed as separate but related to the storage and retrieval system. Each site uses a different approach to managing documents in progress, but in each instance an indexing system other than askSam is used. For example, at ORNL a FoxPro database is used to track documents in process. Authors are officially trained on authoring procedures. A copy of the CDMS system is used to handle controlled documents during the creation process. Three copies of every document in process are maintained; 1 in files, and 2 copies for requests. A policy is in place to officially hand-off documents to the DMC, specifically for archival storage. On the other hand, at K-25 Site, there is no means to track work in progress. There are no clear guidelines, at least none being followed, that control when a document is to be considered complete and therefore handed over to the DMC. There are published procedures for the work flow controlling document creation. Many of the authors, however, do not use them. For example, it is required that every document be passed through a clearance system prior to publication. It is known that some authors and project personnel bypass this step.

4.3.2 The System Infrastructure

ER does not impose an established infrastructure on the solution definition. There is however, an infrastructure in place supporting the current askSam system. Each of the IAs has a 486 or Pentium with a large monitor. The askSam system is at each of the Oak Ridge sites. These are networked together and each DMC has 7–13 nodes on the askSam LAN. askSam 5.11 is running at the satellite DMCs, and migration to askSam 2.0 for Windows, which is what the central DMC is using, is planned for the immediate future.

There are additional infrastructure resources available to this project, but these are discussed in the "Other Site Conditions and Criteria" chapter of this report, since they are not innately tied to this project.

4.4 OTHER SITE CONDITIONS AND CRITERIA

4.4.1 Cultural Issues

It should be pointed out that many of the issues that ER must contend with in this project are not technical or document management-based. There are conditions that are culturally-based that need to be appreciated. For example, project managers are reluctant to turn over their working copies of documents because they insist on near instantaneous access to this information while a project is active. Thus, many
of the ER documents are being held at users’ desks, which accounts for the estimate that only 20% of ER documents are housed in the central DMC. Users are also heavily ingrained in paper as the medium for documents. Despite the production of documents in electronic format (i.e., word processing), business processes have been built around paper. Project managers and other authors need to accept that records management is an issue separate from document creation and that the burden of management rests with the DMC throughout the document’s life cycle.

In several instances the authoring of documents is performed by other subcontractors. This can lead to less control of the process since these subcontractors do not fall under direct ER control. Subcontractors have been known to circumvent rules such as clearance and controlled procedures and sometimes to bypass editors entirely.

There was a sense among some users interviewed that the DMC is the main problem, as opposed to the solution. Some viewed the DMC as a “black hole” into which once a document entered, it was never found again. Others view the DMC strictly as an archival facility. On the other hand, user acceptance of change seemed high. If benefit to the user was demonstrated, migrating the user to a new system for document creation and management would not be a difficult task.

4.4.2 Document Related Issues

Users at all levels, from project managers to technical advisors to DOE auditors, agreed that the focus of their retrieval efforts was on information rather than documents. In essence, not the document itself, but the information contained in the document was required. The accessibility of information was key, not the ability to retrieve an exact copy of the original document. When making a request, individuals are not interested in the document per se, but the information in the document. Although this is a subtle difference, it puts a different perspective on the nature of the application. The focus of the solution should be on making information available (which is in alignment with the CSFs stated for this project), not on complying with a legal requirement to make an image of an original document available.

Several individuals mentioned that often documents are accessed so that pieces of one document can be incorporated into another document. There is a preponderance of sharing common information between documents. Thus, the documents are modular in nature and exhibit a high degree of information overlap.

Security for the documents and the information contained in them is an important feature of this system in that a small percentage of these documents is classified. It is anticipated that the proposed system will manage unclassified nonsensitive information.

4.4.3 The Management Control Information System (MCIS) Infrastructure

If desired, the infrastructure associated with a related application, MCIS, is available to ER IRM. MCIS is a project management system that tracks all programs at the DOE level. This system has a SUN server and a LAN at Y-12, K-25 Site, ORNL, PORTS, and PGDP, which are linked via a TCP/IP and NSF network. Each LAN uses a variety of protocols with Ethernet support to pull the WAN together. There are an average of 60 people per site with access to this network. These individuals represent a superset of the individuals who need access to ER records. The system is operational and was designed for scalability. There is little traffic on the network, and thus, this infrastructure should be viewed as a viable platform on which to deploy an ER records management solution.
5. RECOMMENDATIONS AND OBSERVATIONS

Several key areas need to be addressed in regards to providing an ER solution. Among these are: which documents need to be stored, the format for document storage, the retention schedule of the documents, the authoring process, the structure of the documents, who needs access to the documents, the type of access that is required, and the level of control necessary over the creation and retrieval processes. The best way to address these issues is to consider the technology recommendations made in this chapter of the report. These recommendations are presented in the light of what they will accomplish and their direct benefit to ER.

It is critical for ER to understand that the technology component of the solution is secondary to the need for reengineering and dealing with cultural issues. Although product selection and solution implementation may not begin for several months, the reengineering process and acclimating users to a new work environment can and should begin immediately. (Discussion of the cultural issues that need to be addressed is presented in the “Cultural Issues” subsection of the “Other Site Conditions and Criteria” chapter of this report, and the “Perceived Obstacles” subsection of this chapter.) Additionally, “The Need for Established Procedures” discussion which follows provides insight into the types of procedures that need to be put in place.

5.1 THE NEED FOR ESTABLISHED PROCEDURES

Perhaps the greatest issue that needs to be resolved in this application is the lack of standard procedures and a means to enforce adherence to them. The need to establish procedures that control document creation and handling are as important as the technology used to automate these. Despite a stated objective towards central control over the record creation and management processes, the system is laden with inconsistencies regarding the work flow and protocols for authoring a document, the role of the DMC, the level of integration and timing between the DMC and the authors, the use of keywords, the purpose and definition of a controlled document, the handling of data packages, etc. Although the technologies outlined below will provide a means of control over procedures and redefine certain tasks, the procedures themselves must be clearly defined and an effort made to solicit acceptance and adherence to these procedures. Procedure definition and solicitation of end user acceptance should begin immediately. It is these reengineered procedures that should be automated and facilitated through the deployment of EDMS technology, not the existing procedures. Specific areas that arose during the interviews which warrant ER’s attention include:

- The askSam index—Whether or not the askSam system is used in the future as the document management software for ER, a standard indexing schema that meets the needs of each user should be developed. The satellites have been allowed to customize the index by as much as 25% to meet their site-specific needs. These differences need to be reconciled and supported to the degree necessary to provide intuitive access for each site. In instances where site-specific customization of the index is necessary, this should be accomplished through the addition of the site-specific field to the master template. The system should be configured to provide customized views to the template as warranted by the site-specific needs. Likewise, the DOE auditors expressed concern that the current index does not entirely meet their retrieval needs. As a valid user of this system, their requirements need to be factored into the index design as well. By implementing a master index, the need to reindex records as they are physically transferred from the satellite to the central DMC collection is eliminated. Once developed, use of the index schema should be mandatory for each site. This schema can then be migrated into whatever document management tool is ultimately
chosen for ER. The benefit is that users will have already become acclimated to this index and the requirement to comply. In the case that the design of this index is completed near the time of selection of a new product, the index could be initially implemented in the new system and thus would still save on implementation design time. The only exception to this is the current use of keywords. For reasons highlighted in the full-text retrieval discussion in this report, the continued use of keywords is not recommended, particularly because of the nonstandard approach to their usage and the lack of a controlled vocabulary. Therefore, it is not worth the effort to redesign this facet of the current index. Users should be allowed to use keywords in the manner they are accustomed, until a full-text engine is introduced.

- **Submission of documents to the DMC**—There is no general agreement among various authors as to when and why a document should be submitted to the DMC. Policy should be developed that states the timing required for submissions as well as the role of the DMC. Documents should not be retained by the authors until a project is completed. Users should be instructed that the DMC is not an archival facility but an active document management facility. Although a proactive approach to enforcing these requirements will not be practical until EDMS technology is introduced into the process, user acceptance of these policies can be nurtured in the interim by educating the user as to the benefit of using the DMC as a dynamic resource, and the potential risk of not passing DOE audits if they continue to store documents locally.

Users, including subcontractors and/or agencies external to LMES, should be required to submit their documents to the DMC in electronic format. Despite the use of online authoring systems, users continue to think of the document as a paper-based resource. Required submission of the document electronically will combat this. As outlined in the “Exclusion of Imaging and Retention of Paper” subsection of this chapter of the report, ER may continue to store documents in paper format. Despite this however, it is essential that users begin to think of the document as an electronic resource and accept procedures such as electronic submission. If necessary, the DMC personnel should print the document for long-term storage.

- **Controlled Documents**—Despite the presence of the CDMS, there is little policy concerning the administration of a controlled document. (A CDMS procedure is being drafted to address this.) The decision to control a document is vested solely with the author. While this may be the most logical approach to deciding on the need to control a document, some guidelines should be imposed on the author. Without them, the system can be potentially both over utilized and under utilized. For example, there are occurrences of a document being placed into the controlled document system in spite of the fact that the document is not subject to modification. It seems to be an unnecessary waste of time and effort to control a document that is not subject to revision.

It should be kept in mind that the application of a document management system coupled with electronic delivery of documents may do away with much of the need for controlling a document. (See the discussion on “Document Management with Integrated Full-Text Retrieval” which follows for more detail.)

- **Data Packages**—There is no consistency in the manner in which data packages are handled at each DMC site. Some sites store the electronic version as the file copy, others store the paper. Some maintain an audit trail of any changes to the data in the package, others do not. Clearly, ER should develop a “best approach” to managing these packages and instruct the DMCs and other interested parties on these procedures.
5.2 DOCUMENT MANAGEMENT WITH INTEGRATED FULL-TEXT RETRIEVAL (FTR)

At the heart of the perceived BR solution is a document management system with integrated FTR functionality. The document management system would serve as the DMC, providing indexing capability, security control, and access via the index values and document content.

Under this approach, documents (but not photos and videos) will be stored online in electronic format. Therefore, the documents should be created within this document management environment. This approach will require that authors and/or editors be partially responsible for indexing their documents, but this is a minimal effort. As part of creation, the author would be prompted to enter certain data values (i.e., author name, title, date, and document type) through a data entry screen. It is also possible, depending on the document creation and document management tools used, that these values could be entered automatically without burdening the author.

Unlike the current askSam system which provides the ability to store documents online as well, a document management system will provide extensive security control and revision control. Thus, although a document is not considered final, it could none-the-less be stored and managed by the system. Retrieval could be restricted to certain users until a document is marked “final.” The system’s ability to track revisions would allow authors and editors to check documents in and out, roll back to any former revision, and ensure that retrievals always provide the latest revision. This last feature will be fundamental in reengineering the manner in which controlled documents are handled. Because the system could guarantee that only the latest revision was available, users of controlled documents could be instructed to always refer to the online system when accessing the document. Thus, they would be guaranteed current and accurate information. An update flag could be set in the document record to indicate the date of the last revision which would facilitate a user’s ability to determine how current their perspective on a document is. Through integration with workflow (see the discussion on “Applicability of Workflow” for more detail), individuals on a controlled distribution list could automatically be given notification whenever a new version of a document is made available.

On a related note, if the data packages were stored in the document management system, document management’s revision tracking capabilities could be used to monitor and control updates to this data as well.

Through a distributed document management system, central control could be exercised over a multisite collection. Documents could be stored and indexed on servers local to the various DMC sites. But through a common networked index and central definition of procedures and indexes, the system would in essence be a central DMC. The benefits of storing documents locally are the sense of continued ownership it provides to the end users and the speed of retrieval. It is typically the case that those individuals most interested in a document are physically located at the site where the document was created. By storing these documents locally, document transport across a network would be minimized and thus have a positive affect on retrieval speed. Yet, because the index and access is networked, a user would not have to be concerned with where a document is stored and could, through a single query, access any document no matter where it was stored.

Despite the lack of organization of documents by project in this proposed approach, rapid retrieval of all documents related to a project could still be accomplished by using project name as one of the search criteria.

User access would be enhanced through the availability of FTR as a retrieval resource. Users interested in a particular subject matter could identify relevant documents through a query which would,
in essence, search the content of each document for evidence of this subject in the document. Unlike the
current keyword approach to this type of access, FTR would provide a dynamic assessment of the
documents' contents and would not be subject to an author's or IA's opinion of the value of a document.
A user's ability to query the system would not be bound by the index values of the keywords assigned.
Virtually any information need could be accommodated. Not only would this enhance the retrieval
capabilities of the system, but it would eliminate the time and effort needed to keyword documents.

The combination of FTR with the document management index will allow users to pose queries
in which either or both values are provided. For example, a user could ask for all “correspondence”
written after January 1, 1995, by “John Smith” in which the concept of “contaminated well water” is
presented. Equally as important, however, is the iterative ability of most document management
systems. Users could engage in an online dialog that would allow them, through a series of queries,
to narrow their search results until they have identified exactly what they are looking for. While this
process is in practice available in the current system, it requires printouts and call-backs between the
IA and the user. In the envisioned system, the query process itself will become iterative.

Lastly, a document management system would also automatically maintain an audit trail of
document usage. Thus, a record of who edited and/or retrieved a document throughout its life cycle
will be available. While there was no immediate need for this type of data, several individuals
indicated that this type of information might prove valuable if available.

5.2.1 Benefits Derived

Although some of the benefits associated with the implementation of a document management
system with integrated FTR have been highlighted above, the nature of this component of the solution
is critical enough to warrant an overview of all expected benefits.

- **Records compliance**—Perhaps one of the greatest benefits to come out of this solution
implementation is one that is not directly linked to technology. Benefits will be achieved through
the definition, establishment, and cultural acceptance of clear and concise records management
processes. This will be strengthened through the automation of many, if not all of these processes.
This automation will ultimately result in the enforcement of record keeping and retention
requirements. Assignment of retention periods can be automated and enforcement of retention
schedules can be achieved through record deletion and/or systematic media migration throughout
each record's life cycle. Compliance and quality will be proactively monitored and documented.
The audit trails provided by the document management software and the workflow software will
provide a detailed, accurate, and objective history of each document throughout its life cycle. In this
regard, this benefit is directly linked to the benefit of increased control which is discussed below.

- **Increased control**—This benefit is directly aligned with the first CSF. The audit trails and
sophisticated security of the document management system will allow ER to proactively monitor
usage and content of the documents. If a distributed document management system is used, local
DMCs could remain intact while control is dictated from a central function. The revision tracking
capability of the document management system will also be the foundation to a proactive approach
to controlling documents. The online version will by default always be the most current version of
any document. But, for reasons of historical perspective, former revisions will be available (and
their audit trails) if necessary. With the integration of workflow technology, and/or e-mail, the
system could prompt users regarding the availability of a new revision to a particular document.
• **Enhanced access to information**—This is the most crucial of the benefits and is directly aligned with several of the CSFs. The power of a retrieval engine that couples structured indexing with FTR will result in a system that can meet a wide variety of user needs: from users who know exactly what they are looking for, to new users who have only a vague idea of the subject matter of interest, to experienced individuals who want to apply new perspectives to the information base.

FTR could also be used to compensate for any shortcomings of the structured document index. Queries based on an individual’s name could be executed whether the document was indexed on that name or not. For example, if a document was authored by multiple individuals, and the author field does not track each name, an FTR of the entire document would locate the individual’s name despite its absence in the author index field.

• **Faster access to information**—This benefit is also closely aligned with the CSFs. By storing the documents online, users will no longer have to wait for the IA to copy a document and route it to them. The document will be provided as a direct result of any query.

Documents will also become available for searching more quickly in that they will no longer have to be keyworded and barcoded. Additionally, because the indexing happens as part of document creation, all documents are made available (within security constraints). There is no delay waiting for a document to be submitted and indexed before it can be retrieved.

Because FTR systems typically highlight the terms in a document that reflect the user’s query criteria, users will be quicker in honing in on the portions of the documents that are relevant to their needs. In situations where several documents are retrieved, this ability will increase the speed with which decisions can be made regarding how valuable a document is to the current search.

• **Decreased costs**—The elimination of IA controlled indexing, keywording, bar coding, copying, folders, etc. should result in decreased costs in regards to materials and time.

• **Increased reliability**—The nature of electronic documents makes it virtually impossible for a document to be misfiled or lost. This feature, coupled with increased accessibility and greater control of documents, should lead to a better track record with DOE audits and result in less need for corrective action plans.

• **Facilitated integration**—By storing the data packages online in a document management database, the foundation for integration with OREIS is established.

• **Reusable Information**—Many individuals indicated that the reason they retrieve documents is to reuse information contained in the document. In a paper-based environment this entails rekeying the information. By making the documents available electronically, the cut-and-paste feature of the authoring system can be used to effectively reuse this information.

On this note however, it should be mentioned that migration to electronic media will allow ER to consider using compound document architecture in constructing their documents. These documents would make use of intelligent links not to physically store a common body of text, spreadsheet, or image in a document, but rather to dynamically pull that piece of information into the document from its original source each time the document is accessed. In this way, the latest revision of any of the document’s components would automatically be inserted into the document, keeping the document current. Related to this concept is the use of a compound document manager which is discussed in further detail in the following paragraph.
5.3 HYPERTEXT AND COMPOUND DOCUMENT MANAGEMENT

Within the ER document collection, there is a fair amount of data sharing. Some documents reference other documents. In other cases, pieces of information are replicated from one document to another. In the former case, the application of hypertext would be beneficial. It would allow the cross references made in documents to become not only pointers to additional information but also physical links to that information. In this way retrieval speed on the referenced document would be increased. Furthermore, the ease of navigating through related documents in this way could encourage users to broaden their research by accessing these referenced documents more readily.

The more interesting case, however, is the reuse of information from one document to another. In these instances an authoring approach and document management system that supports the concepts of the compound document should be considered. The more popular example of this is Standard Generalized Markup Language (SGML). In such an environment, the document would be constructed as a series of information components. Components could include bodies of text, spreadsheets, images, and graphics. These information components could be pulled together in a variety of ways, each constituting a document. For example, if a report made references to and included a spreadsheet developed by another user on the system, rather than embedding a static picture of the spreadsheet in the document, a link would be used that would provide the latest revision of the spreadsheet in the document. By utilizing a document management system with support for this type of document architecture, the interrelationships between documents could be proactively monitored. Information components could be locked in certain documents but kept dynamic in others. In this way information integrity as well as document integrity are achieved.

5.4 EXCLUSION OF IMAGING AND RETENTION OF PAPER

Imaging technology is not being recommended as part of the ER solution. Although the document management system envisioned would undoubtedly have the capability to make document images available online through the same interfaces discussed in the “Document Management with Integrated Full-Text Retrieval (FTR)” subsection of this chapter of the report, there is no driving business case to incur the extra expense of imaging. The proposed solution will store desktop publishing (i.e., word processing files) online. These files, although not an exact replica of the “original paper file,” will provide complete access to the document’s content. In a unique showing of unanimous agreement, those interviewed (from LMES management, to technical assistants, to DOE auditors) agreed that the focus of their document retrieval needs was on the information in a document, not in ascertaining a replica of the “original” paper document.

Additionally, there is no benefit in storing images online as a means to backup the paper files since DOE does not accept optical images as the record copy of a document. The volume of documents in this application would necessitate use of optical storage media. (A single page, black and white compressed image requires approximately 50K bytes of storage.) On the other hand, storage of the text files could be met using magnetic storage, a medium accepted by DOE. (A page of text requires approximately 5K of storage.)

A case could be made for document images, if this application focused on providing a legally accepted copy of the document, but, in addition to the optical storage issue, CERCLA and DOE do not accept images as a valid record format. While there is every indication that the federal government will accept document images as valid records sometime in the future, there is no benefit to ER in pioneering an effort in this direction. As stated previously, the focus of this application is on information access not
access to copies of legal documents. The difference is subtle but important. The use of electronic text files as records could save considerable time and money in implementation. Furthermore, text files will transport across the network more efficiently because of their considerably smaller size than images.

Therefore, it is recommended that ER continue to store and manage the currently existing paper files. But, these files will be viewed as archival in the truest sense of the word. In the future, authors would submit their documents to the DMCs in electronic format. The DMC personnel would be responsible for determining the appropriate storage approach and media for archival storage of these files (paper, microfilm, or magnetic). If the document required signatures or other such paper-based handling, the author could be required to submit both versions of the document.

In any case, the paper files should be treated as archived records and moved to a protected area for safe keeping. It is the paper files that will be considered the permanent record, satisfying the 75–100 year retention schedule. The online index should track by a value (such as a box number) where the paper version of a document is stored to facilitate retrieval if a need for the record copy should arise.

5.4.1 Disaster Recovery

It is recognized that this approach does not provide a definite disaster recovery plan. Because the online text documents would not be viewed as the legal record copy of the legacy collection, if the paper documents should be damaged or destroyed, then there would be no legal backup files. ER must make a calculated decision at this point. If ER should meet DOE’s magnetic media storage requirements, then the online version could serve as the record copy and the paper/microfilm copies could serve as the backup. The document management system’s security and audit trails, if properly documented and implemented, would provide a strong legal case for having them accepted as the “best evidence” of a document. If a surer approach is required, then creating a complete duplicate set in paper or microfilm is the safest approach. Of course this approach will bring significant cost. ER must decide the value of this approach and weigh it against the anticipated costs. It must be noted that protection of the legacy collection through a disaster recovery system was not mentioned as a CSF or even as a minor success factor by anyone interviewed. It is, therefore, assumed that such a plan is not viewed as critical to ER, especially if the paper documents are stored in an environmentally protected area.

5.5 APPLICABILITY OF MULTIMEDIA

The online storage and retrieval of the photographs and videos in the ER document collection will require support for multimedia technology. Although most document management systems support these data types, it is the infrastructure issues that make support for multimedia an issue above and beyond the document management approach discussed thus far. Color photographs are a form of imaging. These image files can take up to as much as eight times the storage of black and white text page images (i.e., a 8.5" x 11" color photograph averages 400K bytes of storage). Video is an even more storage intensive media. A five minute video on average requires 300 megabytes of storage. But storage is just one facet that needs to be considered. The necessary scanners and display devices will further catapult the price of the solution. Considering the low volume and less frequent usage of these files, it is not recommended that the initial implementation of document management directly support photographs and videos. The document management system should track these files through an appropriate index (beyond the project box organization used today) but not provide online access. This matter can be revisited after ER has become successfully acclimated to using the document management system for its text documents.
5.6 APPLICABILITY OF WORKFLOW

In the case of ER, workflow’s applicability is primarily focused on the authoring side of the equation. The distribution requirements of controlled documents, however, poses an opportunity for workflow within the records management side of the equation as well.

Because there was no expressed need to shorten the authoring cycle, workflow’s ability to achieve this end will not be the locus of this discussion. Although it is possible through workflow automation to shorten the authoring process, this will be a natural outcome of the implementation as opposed to a targeted goal. The targeted benefit of applying workflow to the authoring process will be in exercising tighter control over the process and ensuring compliance with filing requirements.

By integrating the authoring process within a workflow environment, authors would no longer be confused about the steps necessary to complete a document or be able to bypass steps in the process. Thus, once the process was defined to the workflow system, authors would be required to initiate their writing in this controlled environment. At the completion of each step (i.e., author indicates that she/he has completed the first draft), the workflow engine would automatically transfer the document to the next appropriate step/individual. In this way, each document will go through the proper editing and clearance channels. This system could be loosely integrated with the AR system so that for certain document types, at a certain point in the creation process, a copy of the document would automatically be transferred to the AR system. Ticklers could be used to remind authors of related documents that will be needed within a given time frame. Furthermore, because the workflow system has the ability of keeping an audit trail, AR staff, editors, and other ER Program staff could monitor the document pipeline, predict their incoming workloads, and schedule their time accordingly. Another benefit comes from the fact that by exercising this level of control over the document creation and filing process, the chances of having the electronic versions of the documents accepted as a recovery system copy are greatly enhanced.

As part of the authoring process, the workflow system could be instructed to prompt the author to determine if the document needs to be controlled. If the response is yes, integration with the employee directory would allow the author to determine who requires active notification of the document’s availability and the availability of future revisions. In this way the workflow system would assume the controlling process. The individuals indicated on the control list would receive an e-mail message informing them of the document’s availability and of any future revisions that are filed.

5.7 A PHASED IMPLEMENTATION

In determining an implementation strategy for ER, three separate yet related components need to be addressed: the infrastructure, the order in which technologies should be applied and integrated, and the order in which sites should be addressed.

In regards to infrastructure, it is recommended that ER take full advantage of the infrastructure accessible through the MCIS system. Through MCIS a fully functioning infrastructure consisting of a SUN UNIX server and LAN at each DMC site, including Portsmouth and Paducah, is being made available. Additionally, there is WAN support through TCP/IP. It has been stated that most users who would require access to the ER records system are linked into the MCIS system, and that current network traffic is marginal. Furthermore the system was designed for scalability.

ER should verify if these suppositions are accurate. If so, this represents an opportunity to save time and cost in implementing an ER solution. The network configuration issues have been resolved, and
the equipment acquired. In light of ER's CSF to have a demonstrable system operating within six months, the availability of this infrastructure is a great advantage. Most importantly however, the infrastructure represents a state-of-the-art topology which virtually all document management, FTR, and workflow vendors support.

The deployment of technology issue should be reconciled by addressing the needs manifested in the CSFs. As assessed in this chapter of the report, the technologies proposed will not be universally applied to the ER program, but used in targeted applications. Fig. 2 illustrates where the various EDMS technologies will be applied to ER.

**Fig. 2. Overview of proposed ER system.** This schematic provides an overview of the general approach to document creation, storage, and retrieval within the proposed ER system. A quick comparison between this schematic and the schematic in Fig. 1, which depicts the current system, shows that the two systems are very similar. Despite the occurrence of reengineered processes and the introduction of technology, the ER system will remain the same from this overview perspective. This should facilitate user acceptance. Yet, as indicated in this report, the bottom line results will include radical improvements in document control, retrieval speed and accuracy, and document accessibility. The biggest change presented in this schematic is the elimination of the control subprocess. In this model, this functionality has become automated and integrated into the role of the DMC through the deployment of document management, workflow and e-mail technologies. The boxes drawn around the schematic indicate that document management and FTR are the primary technologies applied to the storage and retrieval side of the equation. Workflow and compound document management are the primary technologies applied to the document creation side of the equation. The overlap between these two boxes indicates that these systems should be integrated.

While control over the document authoring process is necessary, control over the storage and retrieval of the documents is paramount. A solution in the authoring process will not carry the same weight as a solution in the storage and retrieval process. Despite a successful authoring process, if subsequent documents cannot be located or accessed, many of the problems that face ER currently will still not be resolved. Addressing the entire application from cradle to grave might be possible, but would certainly prolong the implementation process. Given the time sensitivity of this application, it would be wise to phase in the technologies along the lines of affected application areas. It is, therefore,
recommended that ER begin by focusing on the management of the documents through a document management system and FTR. For reasons stated earlier, the videos and photographs should not be put online initially.

Following a successful implementation of the document management and FTR solution component, the focus should be shifted to workflow-enabling the authoring process and integrating this environment with the document management environment. At this point, integration with the AR system could be addressed under the guidelines reviewed earlier in this report.

The next step should address the issue of integration with the OREIS and MCIS systems, followed by a determination of the value of providing online access to photographs and videos through multimedia.

Lastly, it is not recommended that ER implement these technologies across all sites simultaneously. The introduction of EDMS technologies and the reengineering of a business process will have serious impact on the organization. User education, acclimation, and acceptance will have to be carefully handled. Historically, the most successful EDMS implementations have been those that have been carefully introduced to an organization a step at a time, rather than one single and massive implementation. In this way, user objections and system problems are handled incrementally which leads to a more manageable implementation.

This being the case, a decision must be reached as to how the ER application should be addressed. Breaking the implementation down into units can be done through the type of document affected or by the users involved. Since there does not seem to be any single or subgroup of documents within this application that warrant special or focused attention, it seems logical to approach the phased implementation through site-specific deployment. In this light, the pilot site recommended is Portsmouth. Portsmouth is being recommended for several reasons:

- Development of a records management system is in the earliest stages. Therefore, there are few legacy issues that need to be resolved.

- The need for a solution is greatest here. Because there is no system in place at Portsmouth, the benefits derived from the proposed solution would have greater impact than on any other site. Although there is room for improvement across DMC sites, Portsmouth stands alone in its recurrent failing of DOE audits. By turning this situation around, the solution would have immediate and radical impact on the site specifically, while demonstrating the power of the solution to the other sites.

- Solution resources are readily available. Portsmouth has allocated funding and staff through an SAIC subcontract dedicated to a pilot effort during FY95.

- Lastly, Portsmouth’s user and document population is smaller than the other sites. Working with this manageable group of documents and end users, the applicability of the solution pilot could be easily demonstrated to the other sites because the needs, roles, and document types are similar across all DMC sites.

After the pilot is successfully completed at Portsmouth, it should be rolled out to the other sites. It is recommended that during pilot development and implementation at Portsmouth other Oak Ridge DMC and/or ER personnel maintain an awareness of and, to what extent possible, participate in system planning and implementation through regularly scheduled project team meetings involving representation from all sites. During this time however, the other sites are encouraged not to begin their own solution implementation. The other DMC sites should use this time to begin resolving cultural issues, establishing
procedures, and improving the quality of the data in the current askSam and CDMS systems. Each site will be subsequently implemented following a successful pilot at Portsmouth. Because the infrastructure is in place, this process should not be time consuming as it is expected that virtually all implementation and design issues will have been encountered during the pilot. Unless a significant change in the ER Program mission occurs, current systems (askSam and CDMS) should not migrate to another interim system until the roll out at all DMCs takes place. Introducing other systems into the DMCs prior to roll out will only complicate the migration path for these sites. Once the document management and retrieval pilot has been successfully installed in each DMC site, energies should be applied to implementing the authoring component of the solution, which primarily utilizes workflow technology. The piloting of this component of the solution should begin with Portsmouth for the same reasons that the storage and retrieval system will be piloted there. Once successful, this application should be rolled out to the other DMCs in a similar manner. Thus, the phased implementation strategy for ER is as follows:

1. Develop standard procedures and policies regarding document creation and storage.

2. Automate this reengineered approach to document management at Portsmouth using a document management system with FTR. Online access to documents should be provided to all concerned parties including a central DMC in Oak Ridge, management, and auditors.

3. Roll out this functionality to the other DMC sites.

4. Automate the authoring process at Portsmouth using workflow technology. This should include integration with the AR system.

5. Roll out this functionality to the other DMC sites.

6. Design an approach to OREIS and MCIS integration—automate if practical.

7. Investigate the need and cost for online access to photographs and videos and automate if practical.

It should be pointed out that one of the biggest benefits in approaching solution implementation in a phased approach such as this is the built-in flexibility. If resources or needs dictate otherwise, steps can be combined, or a single step can be broken down into a series of smaller steps. What should remain constant is the order in which tasks occur. For example, if ER should decide that it needs to automate the storage and retrieval capabilities of all DMCs concurrently, and it determines that it has the resources to do that, then steps two and three can be combined. But executing step six, integration with OREIS and MCIS, should not happen before execution of step two, automation of the storage and retrieval process.

5.8 PERCEIVED OBSTACLES

It is important for ER to proceed with this evaluation realizing that there will likely be obstacles to implementation. Several of these can be anticipated and efforts made to control them from the outset. Among these obstacles, the majority are not technical in nature but cultural. This is not unusual when undertaking an EDMS solution of this magnitude. Because it is anticipated that the MCIS-based infrastructure is established and working, it is expected that there will be little to no problems regarding infrastructure installation and configuration. The main technological issue is that of document conversion.
5.8.1 The Conversion Issue

Although documents created “from this day forward” can be handled in the manner outlined in this report, there is the issue of the legacy documents. If electronic versions of these documents cannot be located on the system, they will have to be converted into electronic format using scanning and Optical Character Recognition (OCR) technology. Pursuant to the design of the system, document images will not be stored, but rather the OCR converted text should be saved in whatever file format is chosen for the ER system (e.g., WordPerfect).

Because the scanning process is not an ongoing project in this application, it is recommended that ER outsource the conversion process. Market research has shown that in instances where the client does not have the need to master the conversion process because of a continued reliance on this form of input, it is economically more sound to contract a conversion expert to perform this service. Depending on the quality of the documents provided to the conversion house, prices ranging from two cents to four dollars per page can be expected. This price however, is offset by the fact that ER would not have to acquire a series of high speed scanners equipped with OCR software.

5.8.2 The Cultural Issues

Many of the cultural obstacles that need to be addressed have already been introduced in this report. Among the more important are:

- The need to establish definitive policies and procedures regarding controlled documents and documents in general, the role of the DMC, and the submission of documents to the DMC. This effort should include acclimating users to new processes, perspectives on the role of the DMC, and new business models.

- Despite the need for radical improvement, user satisfaction is apparently high. This is due predominately to a high degree of complacency. The implementation process must focus on soliciting user support for this system and making the added convenience and benefits apparent to the user. This is especially important in the case of the pilot site, Portsmouth, where the major user concern is lack of space to store paper.

In addition to these issues, on a related note is the fact that ER needs to appreciate the role of a sponsor in this application. Because of three considerations—diversity in the types of individual roles that are being affected, separatist attitudes fostered by widespread geographical location, and general user complacency—it will be necessary to identify a sponsor who can not only advocate the new system but also apply pressure whenever necessary to strongly encourage compliance with the new system. The sponsor will have to make budgets available, either directly or through the cooperation of site managers, and ensure that, once established, personnel will adhere to policy and procedure. As benefits are demonstrated through this phased approach to implementation, the sponsor should solicit increased site support and user champions.

It should be pointed out that this proposed system will only have impact on that part of the process that it can control. The reach and strength of the sponsor will have a direct impact on the success of this project. Particularly important is the solicitation of support from authors and editors. These groups are comprised of individuals from myriad ER user groups: LMES personnel within the ER IRM organization, LMES personnel outside the ER IRM organization, and other third party subcontractors. Exercising appropriate control over the latter two user types is especially critical since they fall outside the direct control of the sponsoring organization.
6. CONCLUSION

In conclusion, it is recommended that ER proceed with the evaluation of document management, FTR, and workflow systems for the purpose of reengineering the ER records management program at all DMC sites. Initial investigation in this area indicates that implementation of these technologies in conjunction with a focused BPR effort can result in increased control over the document creation and management process, improved quality of the documents, increased accessibility of documents, faster retrieval times, and decreased costs.

The next step in this process is a determination of the specific features and functionality required of an integrated document management/FTR/workflow system and a subsequent product examination. This is the subject of the Task III report. In this report, the requirements of the ER project will be used to evaluate and rank candidate product solutions, including six products in use at LMES and/or DOE-OR: ERIS, EDMS, askSam, EiCMS, ProNet, and Viking.
DISTRIBUTION

1. K. M. Andrews
2. L. V. Asplund
3. M. V. Ayers
4. H. L. Boston
5–19. S. L. Burnham
20. J. A. Carr
21. J. S. Colley
22. D. G. Cope
23. K. S. Cox
24. M. F. P. DeLozier
25. W. S. Gibson
26. A. M. Glenn
27. P. L. Goddard
28. R. O. Green
29. D. D. Herr
30. J. A. Hodgins
31. R. K. Holmes
32. L. A. Hook
33. J. P. Hutson
34. B. L. Kimmel
35. R. A. Lawson
36. D. E. Long
37. J. R. Lyons, II
38. L. K. Manning
39. W. C. Maynard
40–42. D. M. Matteo
43. C. W. McGinn
44. L. D. Medley
45. C. L. Mills
46–47. P. T. Owen
48. G. Patterson
49. J. A. Payne
50–52. A. S. Quist (3)
53. L. M. Roseberry
54. B. K. Robinette
55. M. M. Sherrod
56. K. M. Steele
57. H. G. Travis
58. R. K. White
59. T. W. Whitlock
60. M. D. Wiens
61. P. S. Woods
62. Central Research Library
63. ER Document Management Center–RC
64. ORNL Laboratory Records

65. D. M. Carden, DOE Oak Ridge Operations Office, P.O. Box 2001, Oak Ridge, TN 37831-8541
66. I. Montalvo, DOE Oak Ridge Operations Office, P.O. Box 2001, Oak Ridge, TN 37831-8541
68. S. S. Perkins, DOE Oak Ridge Operations Office, P.O. Box 2001, Oak Ridge, TN 37831-8541
69. T. C. Perry, DOE Oak Ridge Operations Office, P.O. Box 2001, Oak Ridge, TN 37831-8541
70. T. L. Brindley, Paducah, KEVIL
71. C. A. Hudson, Paducah, KEVIL
72. G. J. Melendez, Portsmouth, X-3012, MS 7700
73. L. C. Stenzel, Portsmouth, X-3012, MS 7700
74. T. D. Taylor, Portsmouth, X-7725, MS 7614