TECHNICAL REFERENCE DOCUMENT DATABASE SYSTEM (TDOCS) REQUIREMENTS DEFINITION

Prepared for

Nuclear Regulatory Commission
Contract NRC-02-88-005

Prepared by

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

August 1993

logo

TECHNICAL REFERENCE DOCUMENT DATABASE SYSTEM (TDOCS) REQUIREMENTS DEFINITION

Prepared for

Nuclear Regulatory Commission
Contract NRC-02-88-005

Prepared by

Rawley D. Johnson
Joseph H. Cooper
Christopher Moehle
Edward Harloe

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

August 1993

EXECUTIVE SUMMARY

The statutory requirement for the Division of High-Level Waste Management (DHLWM) is to make a construction authorization decision within three years, with a possible one year extension, following the U.S. Department of Energy's (DOE) required submittal of the License Application to the U.S. Nuclear Regulatory Commission (NRC). This requirement will place considerable demands on the DHLWM staff as it conducts its license application and pre-license application reviews. The NRC Overall Review Strategy for the Nuclear Regulatory Commission's High-Level Waste Repository Program (Youngblood, 1993) identifies assumptions and strategies that suggest the need for enhanced computer capabilities and tools to support the staff in performing pre-license application reviews:

- Early availability of preliminary information developed and documented by DOE during the pre-license application phase

- Ability to use the results of pre-license application reviews and supporting investigations

- Ability to streamline the acceptance review process, resulting in compliance reviews that focus less on detailed supporting information and methodologies and more on how the detailed information was used to demonstrate compliance

- Use of compliance reviews to verify the acceptability of DOE's compliance and gain confidence whether DOE's compliance demonstrations are acceptable

- Support of reviews documenting concerns as open items and tracking of these open items

- Development of computer models and codes as an ongoing activity in support of iterative performance assessment

The Advanced Computer Review System (ACRS) is being implemented in the DHLWM to address these assumptions and requirements associated with the NRC review strategy. In March, 1992, the Center for Nuclear Waste Regulatory Analyses (CNWRA) began design and implementation of the computer hardware and software system based on two primary applications: (i) technical computing and (ii) technical document referencing and database access. The technical computing application is being implemented and operation of a client/server-based network with graphics and color display is underway.

A full-text management system, the Technical Reference Document Database System (TDOCS), with imaging of non-textual materials for storage and retrieval of technical reference documents, is being initiated as part of the second application for the ACRS with this report. This application is being pursued in four phases: (i) requirements study; (ii) system design; (iii) system implementation; and (iv) testing, documentation, and training. The options for access of external databases containing primarily DOE technical data are the subject of a related report (Harloe, 1993).

The purpose of this report is to identify overall requirements for TDOCS and facilitate the analysis and decision making necessary to initiate its design. The requirements provided by the DHLWM imply a number of significant needs relative to TDOCS design, maintenance, access and use as follows:

- Compatibility and synchronization with similar capability at the CNWRA

- Incorporation of three in-house technical document databases
- On-demand and routine loading by scanning paper documents and electronic document loading
- Bibliographic header and full-text search
- Facilitation of user confidence in his or her retrieval of all relevant documents and only relevant documents
- Support for staff productivity through the incorporation of references in technical analyses
- Adherence in system design and implementation to NRC policies and standards

The system must meet the specific needs of the DHLWM staff. However, in order to implement a system that is immediately beneficial with the desired functionality, TDOCS should support user confidence in his or her ability to find relevant documents. This concept in a document management system like TDOCS must be understood and appreciated in order to make appropriate decisions and commitments regarding impact, cost, and tradeoffs in the system's usefulness. These decisions and commitments must then be reinforced with policies, procedures, and responsibilities clearly defined by the DHLWM for using TDOCS, so that it will not only provide immediate benefit but evolve with the needs of its users.

Section 2 describes a number of related systems, such as, the Licensing Support System (LSS), DOE's INFOStreams and Improved Records Information System (IRIS), NRC's Nuclear Document System (NUDOCS) and the DHLWM's technical document databases, the ACRS, and the CNWRA's Program Architecture Support System/Program Architecture Database (PASS/PADB). NUDOCS is currently undergoing a revision to handle full-text and images (Johnson and Moehle, 1993); and a modified approach has been proposed for DOE to develop, implement, and operate the LSS as a part of INFOSstreams (Chilk, 1993). All of these systems do or will, when available, manage documents and, thus, may have beneficial aspects useful to defining requirements for TDOCS. The Appendix contains a matrix of requirements derived from a recent analysis of the needs of various NRC Offices for office-controlled document management databases, which was contracted by the Office of Information Resources Management (IRM). The table based on that study and presented in the Appendix measures NUDOCS and a tested prototype system against these requirements. Therefore, the examination of related systems is necessary to identify the unique requirements and pressing policy and compatibility issues that are driven by these systems. These requirements and issues are:

- Immediate functionality and capabilities for the DHLWM with a future interface to the LSS, when it is available, for official high-level waste (HLW) documents
- Header and full-text search capabilities
- Images of non-textual material, such as formulae, equations, tables, graphs, charts, pictures, and photographs
- On-demand scanning to meet immediate needs
- Incorporation of materials and references in analyses
- Connectivity with other systems and databases

The overall set of requirements for TDOCS is analyzed in Section 3 in terms of applicable functions, constraints, and policies. They include the following:

- Database loading: getting documents into the database
- Document processing: cleaning up, entering headers, and indexing full text
- Search and retrieval: accessing the documents

- Document manipulation: facilitating document use
- Administration and maintenance: ensuring system functionality

Based on experience at the CNWRA, it is estimated that the DHLWM might be able to load and process as many as 20 full-text documents per day. At this rate approximately 10,000 documents could be accumulated over a two year period. By comparison, approximately 15,000 documents have been accumulated in the CNWRA library. Also discussed are two important trade-off decisions regarding cleanup after scanning and OCR and the inclusion of bibliographical headers, full-text indexing, storage, and retrieval of page images.

Section 4 concludes the report by summarizing confirmed requirements against proposed and implied requirements. It also provides directions to be taken for resolving policy matters and design issues.

As stated in the CNWRA FY94-95 Operations Plan for the Division of High-Level Waste Management (CNWRA, 1993) estimates regarding scope, cost, and schedule will be updated as necessary during the design phase. Efforts to make TDOCS interfaces completely seamless, provide for easy access to identified document databases, and support imaging functions will be based on identifiable technical alternatives and cost and schedule constraints. The strategy is to provide a system capability that meets the stated requirements and is currently achievable and immediately beneficial. TDOCS will be designed as a small-scale system limited to a relatively low volume of data, and interfaced, first, to the NUDOCS revision and, later, to the LSS, when they are available.

CONTENTS

CONTENTS (CONT'D)

CONTENTS (CONT'D)

FIGURES

TABLES

ACKNOWLEDGMENTS

This report was prepared to document work performed by the Center for Nuclear Waste Regulatory Analyses (CNWRA) for the U.S. Nuclear Regulatory Commission (NRC) under Contract No. NRC-02-88-005. The activities reported here were performed on behalf of the NRC Office of Nuclear Material Safety and Safeguards, Division of High-Level Waste Management. The report is an independent product of the CNWRA and does not necessarily reflect the views or regulatory position of the NRC.

The authors express their appreciation to Ms. Margaret Federline, Dr. Don Chery, Dr. Chuck Interrante, Dr. William Ford, Mr. Robert Carlson, and Ms. Charlotte Abrams for their comments and discussions during the preparation of this report.

ABBREVIATIONS

ACRS	-	Advanced Computer Review System
AMS	-	Advanced Management Systems, Inc.
ANS	-	Auto-dial Network System
API	-	Application Program Interface
ASLBP	-	Atomic Safety Licensing Board Panel
AUTOS	-	Agency Upgrade of Technology for Office Systems
CNWRA	-	Center for Nuclear Waste Regulatory Analyses
COTS	-	Commercially Available, Off-the-Shelf Software
DCD	-	Document Control Desk
DHLWM	-	Division of High-Level Waste Management
DOE	-	Department of Energy
EIS	-	Environmental Impact Statement
GUI	-	Graphical User Interface
HLW	-	High-Level Waste
IRG	-	Integrated Resources Group, Inc.
IRIS	-	Improved Records Information System
IRM	-	Office of Information Resources Management
ISO	-	Information System Organization
LAN	-	Local Area Network
LSS	-	Licensing Support System
LSSARP	-	Licensing Support System Advisory Review Panel
NIH	-	National Institute of Health
NIST	-	National Institute of Science and Technology
NRC	-	Nuclear Regulatory Commission
NRR	-	Office of Nuclear Reactor Regulation
NUDOCS	-	Nuclear Document System
NWPA	-	Nuclear Waste Policy Act
OCR	-	Optical Character Recognition
OCRWM	-	Office of Civilian Radioactive Waste Management
OGC	-	Office of the General Counsel
OIG	-	Office of the Inspector General
OITS	-	Open Item Tracking System
PA	-	Program Architecture
PASS/PADB	-	Program Architecture Support System/Program Architecture Database
PC	-	Personal Computer
QA	-	Quality Assurance
RDBMS	-	Relational Database Management System
RIDS	-	Regulatory Information Distribution System
RM	-	Records Management
SAR	-	Safety Analysis Report
SECY	-	Office of the Secretary
SQL	-	Structured Query Language
SRA	-	Systematic Regulatory Analysis
SRP-UDP	-	Standard Review Plan - Update Project
TCP/IP	-	Transmission Control Protocol/Internet Protocol

ABBREVIATIONS (CONT'D)

TDI	–	Technical Document Index
TDOCS	–	Technical Reference Document Database System
10 CFR PART 60	–	Title 10 of the Code of Federal Regulations Part 60
USGS	–	United States Geological Survey
WAN	–	Wide Area Network
YMPO	–	Yucca Mountain Project Office

1 INTRODUCTION

1.1 BACKGROUND

The statutory requirement for the Division of High-Level Waste Management (DHLWM) is to make a construction authorization decision within three years, with a possible one year extension, following the U.S. Department of Energy's (DOE) required submittal of the License Application to the U.S. Nuclear Regulatory Commission (NRC). This requirement will place considerable demands on the DHLWM staff as it conducts its license application and pre-license application reviews. The NRC Overall Review Strategy for the Nuclear Regulatory Commission's High-Level Waste Repository Program (Youngblood, 1993) identifies assumptions and strategies, including the following, that suggest the need for enhanced computer capabilities and tools to support the staff in performing pre-license application reviews:

- Early availability of preliminary information developed and documented by DOE during the pre-license application phase

- Ability to use the results of pre-license application reviews and supporting investigations

- Ability to streamline the acceptance review process, resulting in compliance reviews that focus less on detailed supporting information and methodologies and focus more on how the detailed information was used to demonstrate compliance

- Use of compliance reviews to verify the acceptability of DOE's compliance and gain confidence that DOE's compliance demonstrations are acceptable

- Support of reviews documenting concerns as open items and tracking of these open items

- Development of computer models and codes as an ongoing activity in support of iterative performance assessment

These assumptions and requirements associated with the NRC review strategy drive the two basic needs that the DHLWM Advanced Computer Review System (ACRS) is intended to address:

- Enhanced technical computing capabilities

- Document management system capabilities

The Center for Nuclear Waste Regulatory Analysis (CNWRA) has participated directly in the analysis, design, and implementation tasks of the ACRS since March 1992. The *DHLWM Advanced Computer Review System Design Summary and Proposed On-Going Support Tasks* report (CNWRA, 1992) recommended ongoing activities to complement and support the ACRS design. The proposed activities were grouped into three primary development tasks and two support tasks. These tasks were discussed with the DHLWM, and, in a management meeting on October 18, 1992, the decision was made to pursue the two highest priority development tasks: design and implementation of a technical reference document database system and access to external technical databases.

The CNWRA has initiated work on both of these tasks, and this is the first report on the task to provide the DHLWM with a technical reference document database system, namely Technical Reference Document Database System (TDOCS). This is a major activity that needs to be pursued immediately to provide the necessary technical document reference and information extraction capability for the DHLWM and the CNWRA technical review staff in the high-level waste (HLW) Program (Meehan, 1993).

1.2 PURPOSE AND SCOPE

The purpose of this report is to identify overall requirements for TDOCS and facilitate analysis and decision-making necessary to initiate the design of TDOCS. This report constitutes an intermediate milestone deliverable on Subtask 6.1 conducted under Center Operations Element, Task 6 (DHLWM) Advanced Computer System for Technical License Review — Support Tasks (CNWRA, 1993). The remaining milestones are system design; system implementation; and testing, documentation, and training.

This report attempts to make explicit all the functions and constraints that need to be supported. In order to meet the needs of the DHLWM, TDOCS must support user confidence and provide value in terms of increased functionality and staff productivity. The major requirements to meet these objectives are:

- Ongoing and routine loading of and access to documents for an up-to-date office database of technical documents
- Accurate and complete document processing
- A variety of search, query, and hypertext access paths and functions to locate documents
- A wide range of practical functions for day-to-day use in staff activities
- Administrative and maintenance functions for a secure and protected system

In short, the system must provide reliable access and practical use of an up-to-date technical document database.

Based on experience at the CNWRA, it is estimated that the DHLWM might be able to load and process as many as 20 full-text documents per day. At this rate approximately 10,000 documents could be accumulated over a two year period. By comparison, approximately 15,000 documents have been accumulated in the CNWRA library. Also discussed are two important trade-off decisions regarding cleanup after scanning and Optical Character Recognition (OCR) and the inclusion of bibliographical headers, full-text indexing, storage, and retrieval of page images.

As stated in the CNWRA FY94-95 Operations Plan for the Division of High-Level Waste Management (CNWRA, 1993), estimates regarding scope, cost, and schedule will be updated during the design phase. Efforts to make TDOCS interfaces completely seamless, provide for easy access to all document databases, and support imaging functions will be based on identifiable technical alternatives and cost and schedule constraints. These efforts will be fully discussed in a follow-on design document. The strategy is to provide a system capability that meets stated requirements and is currently achievable and immediately beneficial. TDOCS will be designed as a small scale system, limited to a relatively low volume of data, with immediate but limited benefits. It will interface to, and complement, both the Nuclear Document System (NUDOCS) revision and the Licensing Support System (LSS), when those systems are implemented, providing ongoing and significant support for DHLWM users. NUDOCS is currently undergoing a revision to handle full-text and images (Johnson and Moehle, 1993); and a modified approach has been proposed for DOE to develop, implement, and operate the LSS as a part of INFOStreams (Chilk, 1993).

1.3 DESIGN AND DEVELOPMENT REQUIREMENTS

The design and development requirements provided by the DHLWM for TDOCS are driven by two major requirements and a number of other considerations. The first requirement concerns the capability for on-demand text and image scanning, OCR, indexing, storage, full-text search, and retrieval of selected documents. The second requirement concerns the capability for routine scanning and loading of a selected set of documents that are submitted to the NRC and are relevant to the HLW program for full-text search and retrieval. Other considerations include (i) implementation on the ACRS server, using as much commercially available software as possible; (ii) the consolidation of three in-house technical databases; (iii) the incorporation of text and images, including color, in technical analyses; (iv) access to documents in other databases, such as the NRC's NUDOCS and the CNWRA's Program Architecture Support System/Program Architecture Database (PASS/PADB), and transfer of individual documents and those in technical data packages from DOE's INFOStreams; and (v) dial-in to on-line library bibliographical services with access and download capabilities.

The requirements provided by the DHLWM imply a number of significant needs related to TDOCS design, maintenance, access, and use. These needs are as follows:

- *Compatibility and synchronization*. Because of the close relationships between the DHLWM and the CNWRA and their need to share documents and data, and because the plan is to implement TDOCS at both sites, it will be important for the design to be compatible with systems at the CNWRA in order to synchronize TDOCS at each site.

- *On-demand and routine loading*. On-demand and routine loading will need to include not only the scanning of paper documents but also the loading of electronic documents, including such sources as on-line, tape, optical disk, floppy disk, and others.

- *Policies, procedures, and responsibilities*. On-demand and routine scanning and loading of documents and the synchronization of databases at both sites will require specification of policies, procedures, and responsibilities for the proper maintenance of TDOCS.

- *Structured queries and full-text search*. It is likely that the DHLWM staff will need to be able to access documents through both structured query and full-text search — that is, it is expected that they will, in some cases, need to locate documents by specific title, author, keywords, and other header fields, while in other cases, they will need to locate all documents that refer to a given concept through full-text search.

- *User Confidence*. TDOCS should support user confidence in his or her ability to achieve recall, the capability of finding all relevant documents, and precision, the capability of finding only relevant documents.

- *User productivity.* Use of the system to support technical analyses implies that there are needs for viewing, printing, and transferring documents; cutting and pasting, storing references, embedding notes, and creating hypertext links.

- *Adherence to standards.* System design and implementation must adhere to all relevant NRC policies and standards.

1.4 USER CONFIDENCE AND PRODUCTIVITY

The system must meet the specific needs of the DHLWM staff. However, in order to implement a system that is immediately useful, system requirements must be driven by the notion of user confidence. That is, TDOCS must support user confidence in all of its major functional areas:

- Database loading
- Document processing
- Search and retrieval
- Document manipulation
- Administration and maintenance

These five functional areas are diagrammed in Figure 1-1, which illustrates the movement of documents through the system and suggests the key role of policies, procedures, and responsibilities of administration and maintenance. An important implication of the diagram is that TDOCS software can only support confidence. Confidence in the system will be gained only by the adoption and practice of policies, procedures, and responsibilities for it use. Confidence can be maximized at the loading stage through policies for the appropriate selection of:

- Documents from in-house databases
- Incoming technical documents
- Documents from other databases and on-line services

Confidence can be maximized at the processing stage through the appropriate procedures for:

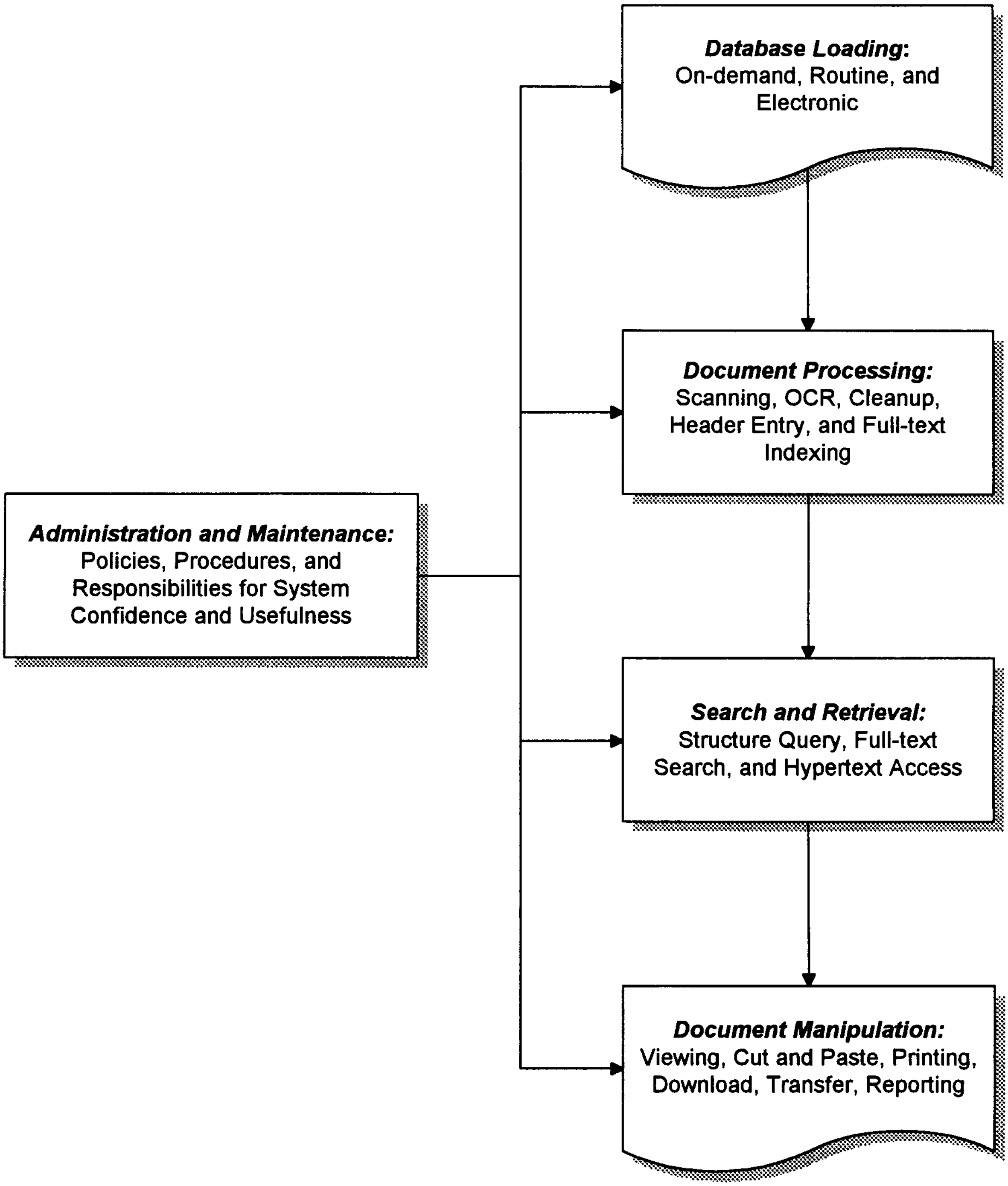
- Document scanning, OCR, and accurate and complete cleanup
- Accurate and complete bibliographic headers

Confidence can be maximized at the search and retrieval stage by providing these appropriate access tools:

- Full-text search on all documents
- Structured query search on headers
- Hypertext access to documents

Confidence can be maximized at the document manipulation stage by providing user functionality for:

- Cut and paste to enable users to incorporate portions of documents directly into analyses
- Document printing, downloading from the server, and transfer via electronic mail (e-mail)


***Database Loading*:**
On-demand, Routine, and Electronic
Document Processing:
Scanning, OCR, Cleanup, Header Entry, and Full-text Indexing
Administration and Maintenance:
Policies, Procedures, and Responsibilities for System Confidence and Usefulness
Search and Retrieval:
Structure Query, Full-text Search, and Hypertext Access
Document Manipulation:
Viewing, Cut and Paste, Printing, Download, Transfer, Reporting

Confidence can be maximized at the administration and maintenance responsibilities stage by:

- Protecting documents with user privileges
- Setting up appropriate user accounts
- Providing standard password protection and system backup

The notion of user confidence in document management systems like TDOCS must be understood and appreciated in order to make appropriate decisions and commitments regarding its impact in terms of cost and tradeoffs in the system's usefulness. These decisions and commitments must then be reinforced with policies, procedures, and responsibilities by the DHLWM for using TDOCS so that it will not only provide immediate benefit but evolve with the needs of its users.

Immediately following are major sections that review existing and planned systems; specify, analyze, and define the scope of requirements; and draw conclusions on important requirements. The status of each of these existing and planned systems of the major organizations involved is reviewed in Section 2 to identify specific capabilities, interface requirements, schedules, and other considerations necessary for effective TDOCS implementation. Detailed requirements for the design of TDOCS, as derived from the NRC and the CNWRA systems, as well as discussions with the DHLWM, are defined in Section 3. In Section 4, conclusions are drawn concerning system requirements and design confirmed by this report. Other issues that remain to be determined during design are formulated as subsequent action items.

2 RELATED SYSTEM CONSIDERATIONS

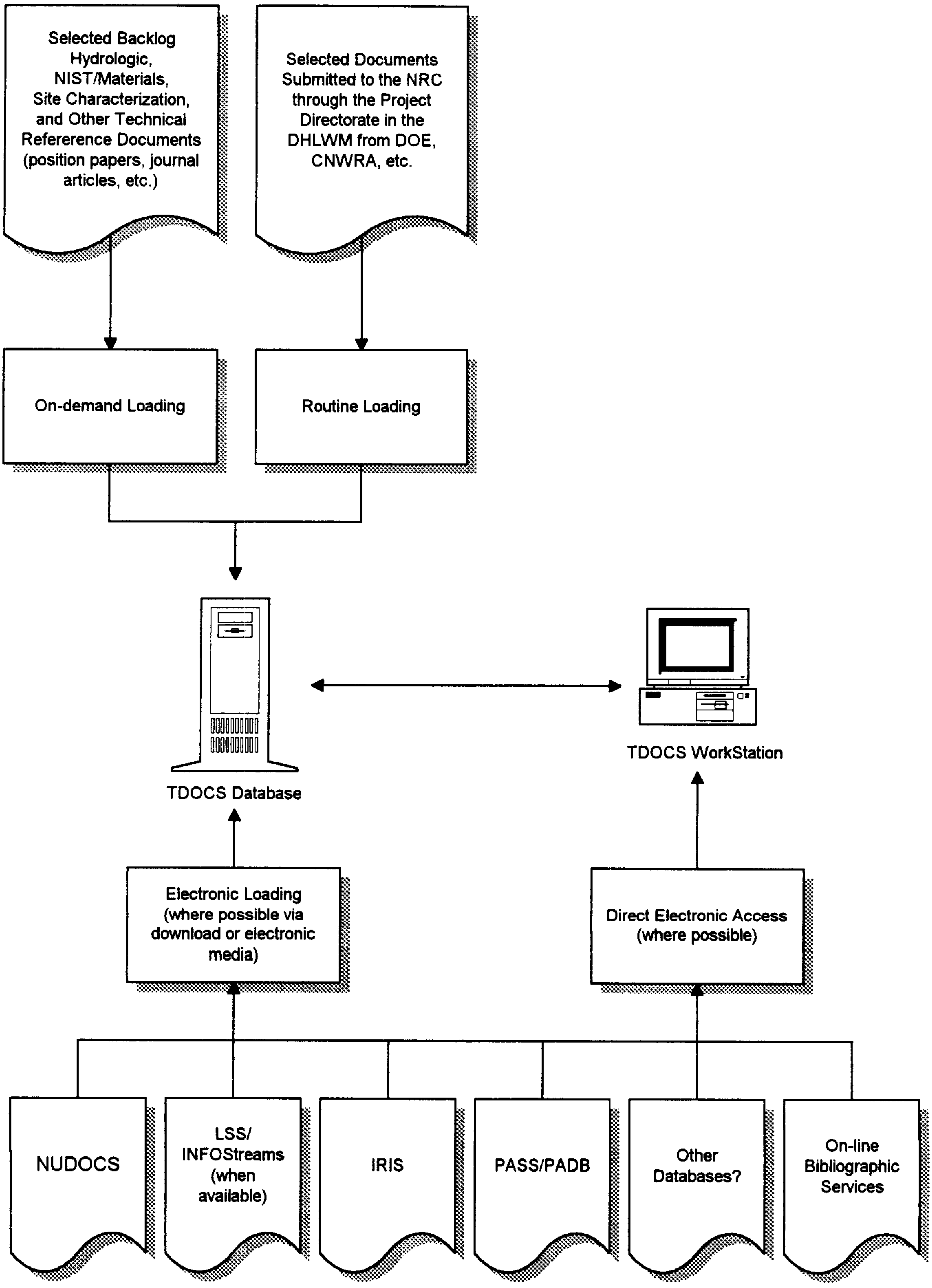
TDOCS capabilities must be designed in a useful and beneficial way. Such a design requires an understanding of the DHLWM manual and automated operations in order for the system to conform with current and planned operations and to meet the needs of its intended users. It is also reasonable, when developing a currently achievable system, to investigate other operations and systems being used, developed, or planned to provide similar capabilities. The purpose of this section, then, is to establish a set of related requirements applicable to the design and implementation of TDOCS.

The LSS will be discussed, since there was a negotiated rulemaking for the establishment of the LSS to contain all the documentary material for all parties to the HLW licensing process. The recent decision by the NRC regarding the long-range plan for the LSS is described. The DOE's planned capabilities with INFOStreams and records management for licensing are presented as stated at the Third Annual International Conference on High-Level Radioactive Waste Management. It is estimated the DOE will produce over 85 percent of the HLW program documents, which the NRC will review in a selected manner. Potential NRC office-controlled text management systems are surveyed. This survey includes a requirements analysis conducted by Advanced Management Systems, Inc. (AMS) for the NRC whereby appropriate staff at the NRC were interviewed to document Agency plans to revise NUDOCS and meet the document management needs of some of its offices. The ACRS being designed and implemented by the CNWRA and the DHLWM is reviewed. TDOCS is planned as a major application of the ACRS. The full-text management requirements proposed by the CNWRA for PASS/PADB are surveyed. This system will be used as part of the overall relational database system for preparing, searching, and retrieving Systematic Regulatory Analysis (SRA) records.

The requirements derived from discussion of the above systems are summarized at the end of this section. Figure 2-1 illustrates the essential sources of technical documents for TDOCS. The diagram depicts a generalized system, the details of which are discussed in this and subsequent sections. It is important to point out that on-demand loading focuses on individual documents, including selected backlog technical references in-house databases, position papers, journal articles, and so on; while routine loading focuses on document types, including selected document types submitted to the NRC and passing through the Project Directorate from the DOE, CNWRA, and other participants. Some systems, particularly the NRC prototype and office-controlled systems, are discussed in this section for their relevant requirements rather than for a need to interface with them and, thus, are absent from the diagram. The presence of some other systems in the diagram does not imply a requirement for an immediate interface to them. The strategy is to provide a system capability that is currently achievable and immediately beneficial. While it is possible to plan for future expansion, it is not possible to design interfaces to systems, such as the NUDOCS revision and the LSS/INFOStreams implementation, that do not yet exist and whose database design, communications protocols, and implementation schedules are not yet formulated. TDOCS will be designed as a small-scale system limited to a relatively low volume of data, and interfaced, first, to the NUDOCS revision and, later, to LSS/INFOStreams, when they are available.

2.1 THE LICENSING SUPPORT SYSTEM

The LSS, an electronic information management system, will contain all the documentary material of all parties to the licensing process for the HLW repository. It is expected to provide access to full text and images with technical data arranged in data packages. The NRC has been examining ways to have an effective LSS, reduce its overall cost, and achieve a more workable alignment of LSS


Selected Backlog Hydrologic, NIST/Materials, Site Characterization, and Other Technical Refererence Documents (position papers, journal articles, etc.)
Selected Documents Submitted to the NRC through the Project Directorate in the DHLWM from DOE, CNWRA, etc.
On-demand Loading
Routine Loading
TDOCS Database
TDOCS WorkStation
Electronic Loading (where possible via download or electronic media)
Direct Electronic Access (where possible)
NUDOCS
LSS/ INFOStreams (when available)
IRIS
PASS/PADB
Other Databases?
On-line Bibliographic Services

responsibilities between the NRC and the DOE. Recently, the Commission directed the staff to pursue discussions with DOE and the Licensing Support System Advisory Review Panel (LSSARP) on a modified approach for the development and operation of the LSS. The modified approach would have DOE develop, implement, and operate the LSS as part of INFOStreams, with the NRC providing oversight of its development and operation, and compliance assessments and audits of the LSS contents. The plan is to have the LSS loaded and accessible by all parties one year prior to license submittal (Chilk, 1993).

It is currently anticipated that the LSS will not be on-line until the year 2001. Even then it will not provide access to the DHLWM documents or in-house databases. TDOCS is needed now to support control and retrieval of documents by the DHLWM staff for technical reviews in the pre-licensing stage (Youngblood, 1993). The DHLWM will need to acquire individual documents and those in data packages from the DOE, in particular, before the LSS is on-line, and download them for local access. The longer-term goal for the LSS to support the retrieval of all HLW program documents does not mean that the TDOCS will go away when the LSS is available. TDOCS will still be needed for control of DHLWM documents at a scope and level beyond that which the LSS will provide. Nevertheless, LSS/INFOStreams development must be monitored to ensure the compatibility of TDOCS with them for support of submittals to the LSS when it is available.

2.1.1 Licensing Support System Technical Data Packages

The CNWRA (Johnson et al., 1991) investigated alternative ways of making LSS packaged documentary materials available. Estimates are that up to 50 percent of the documentary material related to licensing a HLW repository is unlikely to be suitable for entry into LSS in the form of searchable text. Some of these materials (i.e., graphic and handwritten materials) can be scanned as images for viewing on LSS screens; the rest consists mainly of magnetic material, usable only by means other than electronic display. Since full-text search techniques cannot be used to find information in these documentary materials, it will be necessary to rely on the adequacy of bibliographic headers prepared to describe them.

The governing LSS Rule, while providing for packaging of documentary material, does not indicate how that material should be identified aside from the submittal of a bibliographic header based on a package's table of contents. The CNWRA recommended that individual, supplementary headers be assigned for finished products and machine-dependent items. Commentary and graphic/handwritten material are adequately described in a package's table of contents. If TDOCS is to be able to handle packages, then it must accommodate headers at these levels. Furthermore, it is not clear how the LSS will actually provide for packaging of documentary material, since the NRC has recently decided to pursue an approach whereby the LSS is intended to be developed and operated as a part of the DOE's INFOStreams (Chilk, 1993). The initial focus of TDOCS will be on individual documents; but, as the packaging of documentary materials becomes more defined for the LSS, those packaging concepts may later be incorporated in TDOCS.

2.1.2 Information Horizon and Hypertext

The CNWRA (Johnson et al., 1991) discussed a number of interesting document management concepts that the LSS could employ to meet requirements for access to packages, at least two of which have direct relevance if TDOCS is to be able to handle packages. These concepts have to do with the notions of information horizon and hypertext.

The information horizon is the point beyond which information cannot be effectively found. Factors that affect the information horizon include access protocols, indexing, retrieval efficiency, relational linkages, degree of specificity of search parameters, and so on. To deal with this problem in the LSS, indexing is intended to have two major thrusts: bibliographic search and full-text search. Bibliographic headers allow users to access information through fixed fields, much like using a library card catalog. With bibliographic headers, the information horizon is limited by such factors as the content, accuracy, completeness, and quality of the headers. Policies and procedures controlling these factors are an important consideration.

Another method to dramatically expand the information horizon is full-text indexing. This method allows users to search for individual words, combinations of words, phrases, and so on. However, if the implementation of full-text search and retrieval is not fairly sophisticated, search results can overwhelm the user, obscure the desired information, and render it beyond his or her effective information horizon. In addition, full-text search can be applied only to textual documents, not images.

Therefore, a balanced approach is required, one that employs bibliographic headers for searching the information base for titles, authors, and the like; and full-text search to check for information not fully contained in bibliographic headers, namely the full text of each document.

The term, "hypertext" is used to describe a relatively new information storage and retrieval technology. This technology permits logical relationships between different text and/or graphical records to be embedded within the data. Hypertext permits users to navigate quickly and efficiently between related units of information without having to deal with conventional commands, queries, menus, and selection lists. In a hypertext application, the logical relationships, called hypertext links or "hyperlinks," are constructed such that certain words and/or phrases in the textual data are internally associated with other records, images, or processes. Typically, the textual information is presented to the user on a display screen with certain words or phrases highlighted. These highlighted words or phrases are the hypertext links. The user, using a mouse or other pointing device, identifies and selects one of the highlighted hypertext links. The system responds by activating the hypertext link and executing the procedure associated with the link so that the related textual data or image is retrieved and displayed.

A fairly straightforward application of hypertext may be illustrated by the example of the table of contents and the index of a book. In a typical table of contents, one finds a list of chapters or topics that are arranged sequentially to represent the physical organization of the book. In the index, one finds a list of key words arranged alphabetically to permit access to the book through concepts. Thus, the table of contents and the index represent alternate methods of accessing information contained in the book. Page numbers are associated with each entry in both the table of contents and the index to tell where the beginning of the desired text may be found. So when one wants to find some information in a book, the typical approach is to open it to either the table of contents or the index, find the desired entry, and then turn to the appropriate page. In other words, a logical relationship is established for each entry in the table of contents or index that points to the corresponding text for a particular chapter or concept. The page number is the vehicle used for traversing that logical relationship.

In a hypertext implementation of the table of contents or index, the same logical relationships would exist. However, the linkages between the table of contents and the chapters, or the index and the text addressing certain concepts, would be embedded in the table of contents and the index, respectively. A hypertext table of contents or index would not need to show the page numbers. By "pointing to" and selecting the desired entry in the table of contents or index, the user would be able to select and display the desired text immediately. In effect, the hypertext application would automatically "turn the pages" so that the first page of text for a selected table of contents or index entry would appear immediately.

Hypertext is developed by applying computerized links to a document in order to establish logical relationships among the document's words, phrases, diagrams, and images. A document display application capable of interpreting these hyperlinks can then be used to highlight document links. When a user selects a link, the application follows the link to related information, retrieves, and displays it. Hypertext applications can be straightforward or complex, depending on the nature of information relationships and user requirements. It has been suggested that this technique could be applied to the table of contents of LSS packages. Each package of documentary materials includes a table of contents containing an entry for each line-item unit of information in the package. These entries should include a short description of the item, perhaps as much as would be entered in a bibliographic record. Thus, tables of contents could be captured and converted to text format. As such, full-text search could be applied to the converted tables of contents, and individual items in the table could act as hypertext "entry points" into the related documents. The same principles could be applied to any large document, for example 10 CFR Part 60.

2.2 DEPARTMENT OF ENERGY AND THE INFOSTREAMS

The following information from DOE's session on "Information Management of the U.S. HLW Program" is excerpted from the *Proceedings of the Third Annual International Conference on High-Level Radioactive Waste Management*. The two articles referenced in this section do not necessarily reflect the current views of the NRC or the DOE. There will no doubt be some adjustments to these DOE plans with the forthcoming meeting of all parties with the LSSARP on the recently modified approach to the LSS discussed in the previous section (Hoyle, 1993).

2.2.1 INFOStreams

The information in this section on INFOStreams is excerpted from a paper presented at the aforementioned session, titled "Information Management for the Department of Energy Office of Civilian Radioactive Waste Management" (Cerny, 1992).

> The major goal of INFOStreams is to logically organize program information into streams that can be processed, stored, accessed, and disseminated to accomplish Office of Civilian Radioactive Waste Management (OCRWM) goals and objectives. Users are linked through personal computers into local and wide area networks and into mainframe capabilities. Groupware will enable them to create, route, concur on, and track documents, while an audit trail is being automatically generated. Different forms of information are treated appropriately. Management information systems and baseline technical data systems, for example, are standardized within a program wide database structure. Specialized analytic programs and models, however, would have only network access through appropriate interfaces.
>
> Information will be routed into streams both manually and with the assistance of a rule based expert system. The use of this expert system represents one of the solutions offered by technology. A major breakthrough results from the paradigm shift of moving from the management of information, documents, and records to the

management of information *based on the value of the content of the documents to the program*. As information is created or input into the system, a bit mapped image is produced as well as ASCII text, if appropriate, and the document is channeled by content into the appropriate stream. At the same time, the content will determine its ultimate disposition. Will it be put into a computerized database and in what form? For what period will it be kept up? Is it permanent information that must be generated on microfilm and sent to the National Archives? One of the reasons that the U.S. is losing its institutional memory is that disposition of official records in a paper based world are governed by broad categories. Not until rules can be applied concurrently with the generation of the information, as in INFOStreams, can this process be substantially sharpened.

The INFOStreams concept is a natural outgrowth of a major information management initiative. Several years ago, DOE and the NRC conceived of the idea of the LSS. This $200 million computer system would serve as the sole source of document discovery for all parties to the license hearing for the geologic repository. Information that is relevant to, or could lead to relevant information, would be included. All parties would enter their information in exchange for use of the system. A successful negotiated rulemaking among the potentially affected parties on the functions, use and administration of the LSS was held in 1988-1989.

As we began to develop the LSS concept, it became clear that the intake process for the LSS could not be implemented out of context of the Office of Civilian Radioactive Waste Management (OCRWM) program, as the rule envisioned. Since DOE has 85-90% of the information that will reside in the LSS, it must be captured, screened, and indexed following stringent procedures that include quality assurance. The Audit trail, the pedigree of the information, and the defensibility of what was collected, and how, are parameters that must be overlaid upon the technology used for the physical processing. While the license application is developed as one "stream," with pointers to supporting data, ultimately the totality discovery during the hearing is yet but another "stream."

2.2.2 Department of Energy Records Management

The information in this section on INFOStreams is excerpted from a paper presented at the aforementioned session, titled "Records Management in Support of the Licensing Process for the High-Level Radioactive Waste Facility" (Sheats, 1992).

At the Yucca Mountain Project Office (YMPO) the type of information that will be required to support the licensing effort will be contained in a wide variety of media. The media will range from the hard copy document to imaging (both microform and electronic). The one critical element throughout the anticipated media is that information is sufficiently recorded and can be transferred to others for consideration. The transferring or distribution of this recorded information will be the responsibilities of three separate entities. They are:

- The Originator of recorded information (Source)

- The Information System Organization (ISO)
- The Records Management (RM) organization

The source will use the tools developed by the ISO [Computers, Local Area Networks (LAN), Wide Area Networks (WAN), etc.] to generate or record the required information. The ISO generally will be responsible for how the tools are integrated as a coherent distribution system. RM will provide the critical link to that system designated to function as the single source of licensing information, in this instance the LSS.

The information distribution scheme positions the RM organization in a unique role that is not performed in an archival operation. The archival operation consists of receiving recorded information after it has exhausted its useful life cycle. An example of this is the retrieval of calculations or drawings. Seldom is RM called upon to provide current information (latest revision). The source of current information is usually the domain of Document Control. The RM only receives the "current information" after it has been revised upward. It is then called on to provide information for historical or legal purposes.

In the LSS distribution scheme, RM will be called on to insure the maintenance of information in a current state. This function is not unlike that of Document Control and firmly establishes RM as a "Pro-Active" member of the information distribution team.

As the logical entry point to the LSS and in order to perform effectively, RM must understand the licensing process and the information requirements imposed by that process.

RM AND THE LSS

The Nuclear Waste Policy Act (NWPA) mandates a three (3) year license proceeding for a HLW repository. This period is a significant reduction in time from the 6-13 years to license power reactors. This reduction complicated an already delicate process. Further complicating the license process is that the NRC has never licensed a repository. Experience has shown that a good portion of time in any hearing process, either judicial or administrative, is consumed with legal document discovery and motion practice. This will be particularly true in licensing the repository. The immense volume of information that will be produced and require evaluation requires special consideration.

In order to facilitate both the time line for licensing and the anticipated volume of documentation, Title 10 of the Code of Federal Regulation (10 CFR) Subpart J sets forth the procedures that are to be followed by all interested parties. Pursuant to Subpart J all relevant information from all parties to the hearing would be computerized and made available in both full text and image form. The design and development of this computerized system called the LSS is the responsibility of the DOE.

The primary purpose of the LSS is to support an expedited licensing process. To carry out the intended purpose and withstand the rigors of legal challenge the LSS must contain all recorded information (documentation) required to support the repository licensing process. This will be accomplished by providing electronic access to documentary material and electronic transmissions of filing, order and decisions. Documentary material is that information that is relevant to the licensing of the site. To determine relevancy, Integrated Resources Group, Inc. (IRG) developed and is currently testing an inclusion/exclusions list to be used by RM. This inclusion/exclusion list will determine what records, that pass through the RM process, are to be designated as LSS records.

LICENSING DOCUMENTATION

The documentation required to support the licensing of the repository can be found in 10 CFR 60. The attached figure (Figure 2-2) shows the licensing process, the general products resulting from the various sub-processes and the 10 CFR 60 controls over each of those sub-processes. As categorized by IRG, the information that will be migrated to the LSS has four (4) levels as follows:

- **Level 1: Primary Information**

Technical (published) reports, procedures, correspondence, etc.

- **Level 2: Supporting Information**

Scientific notebook, related correspondence, comments directly applicable to data.

- **Level 3: Processed Information**

Processed data, electronic data, raw data, circulated drafts, lab notes.

- **Level 4: Preliminary Information**

Raw data used to bound an experiment, preliminary data not directly applicable, extra computer runs not utilized, preliminary drafts.

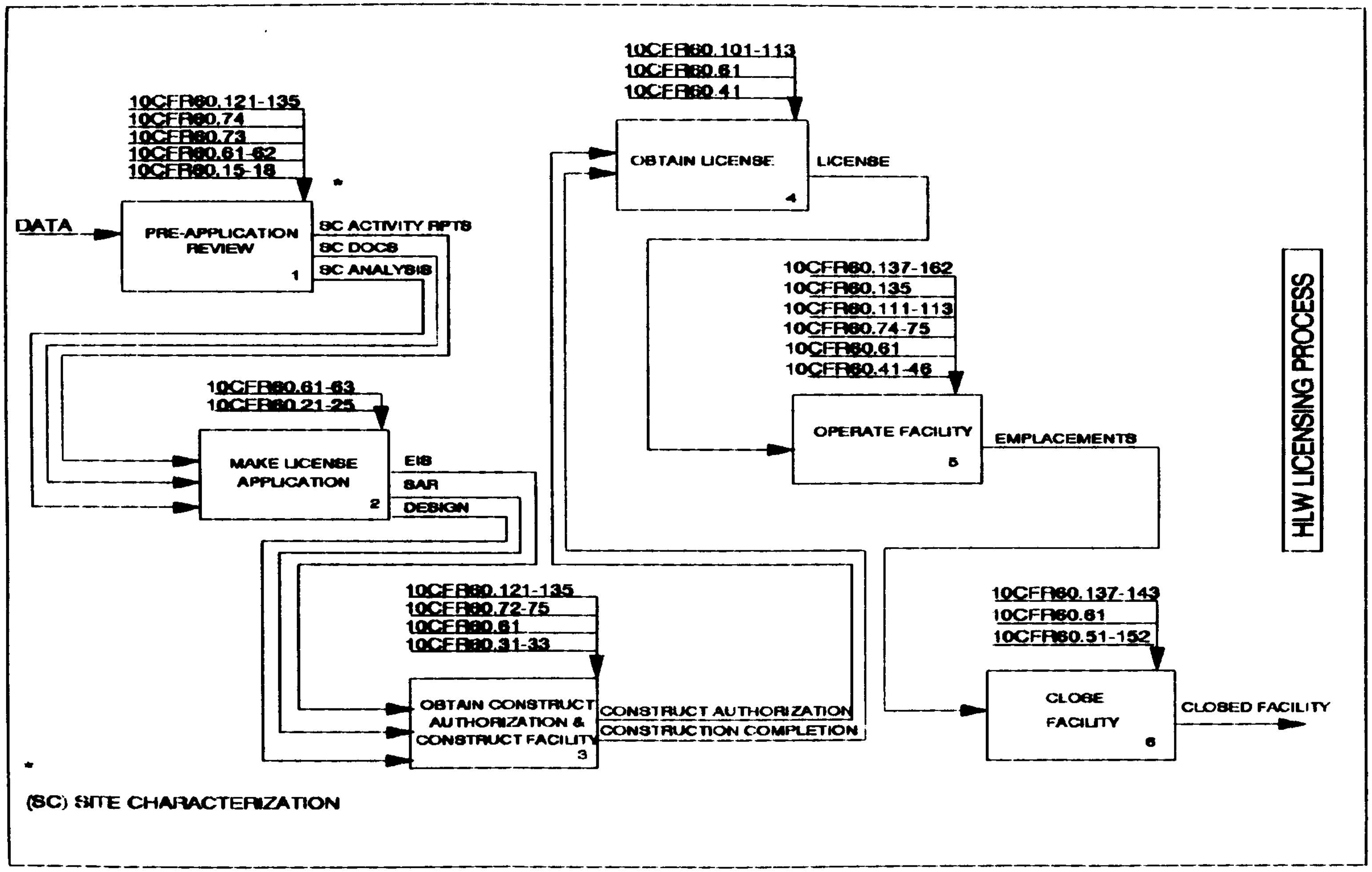
In this hierarchy only level 4 can be considered irrelevant or non-applicable to the Licensing process.

This documentation will be produced from the following activities and may be considered "Licensing Basis Documents."

- **Site Characterization Plan:**

Although not specifically a licensing document, it provides the framework for those activities that will produce information required for the license application (See 10 CFR 60.18).

RECORDS MANAGEMENT/LICENSING PROCESS



- **License Application:**

The license application is the key licensing document (See 10 CFR 60.21). It will consist of General Information, the Safety Analysis Report (SAR) and the Environmental Impact Statement (EIS).

Information required to support the SAR is comprised of that which describes, analyzes and provides information for subsequent discussions. These discussions will involve assessments of the site. This will include identification and evaluation of the geologic setting and design of both the surface and subsurface operating areas.

The evaluation of operating area designs will focus on performance requirement of the structures, systems and components that are considered important to safety. This requirement also includes those structures, systems and components that require research and development. These requirements do not differ greatly from those for the licensing of a nuclear power plant (see 10 CFR 50.34).

Additional requirements are the details of the Quality Assurance Health Physics, and Emergency Plans and Programs. Again these differ only marginally from those for nuclear power plants.

2.3 NUCLEAR REGULATORY COMMISSION EFFORTS TO MEET DOCUMENT MANAGEMENT NEEDS

This section examines the NRC NUDOCS and current efforts to meet document management needs in the various offices of the NRC. First, the system and procedures for document submission, search, and retrieval are briefly described. Next, plans to revise the system are examined. Finally, a series of efforts by the NRC to define similar requirements, to design a prototype, and to produce a distributed, customer-controlled, document management systems are surveyed.

2.3.1 Nuclear Document System

The official repository of documents at the NRC is the NUDOCS. It is the repository for over 2 million records of documents, some full text, some abstracted, that have been generated or received by the NRC since 1978. Effective license application review requires access to these references for incorporation into analyses and written reports. What follows is a brief description of procedures for document submission, search, and retrieval; it is a paraphrase of a more detailed description (Youngblood, 1992).

> DHLWM staff are responsible for the placement of internally and externally generated documents into the document control system. Internally generated documents must be submitted according to procedural guidelines to the Document Control Desk (DCD) for processing, distribution, and submission to NUDOCS by secretarial staff. The DCD is principally responsible for ensuring that externally generated documents are distributed to DHLWM staff and submitted to NUDOCS; however, the DHLWM staff are responsible for returning hand-carried documents and those marked "Personal" or "Addressee Only" for processing.

NUDOCS staff are responsible for processing submitted documents. For each document, they prepare a microfilm copy and a bibliographic header with keywords for the NUDOCS database. The microfilm is distributed to all NRC NUDOCS WorkStations. Bibliographic headers are inserted into the NUDOCS database, an Oracle database running on a Data General MV/40000. Some documents, many related to HLW, have been designated for full-text processing; in fact, electronic versions are also submitted with paper copies. All documents except those containing proprietary or safeguards information are made available to the public.

Once a document has been submitted to NUDOCS, it is available for search and retrieval at a NUDOCS WorkStation. DHLWM staff can access NUDOCS over the Agency Upgrade of Technology for Office Systems (AUTOS), specifically using "SmarTerm" software over an Auto-dial Network System (ANS). On line, the interface allows users to search by menu or command, browse sets of records, or employ predefined queries. They can perform structured queries on headers or full-text search on available scanned documents to locate bibliographic references. These references provide accession numbers that staff can use to locate microfilm copies of documents or download available full-text documents. In the case of microfilm copies, staff must key in the text in order to incorporate technical references into reviews.

In discussions with them (Johnson and Moehle, 1993) DHLWM staff have expressed concern with NUDOCS as a document search and retrieval system. These concerns are based on several factors. First, turnaround time, between document submission and its on-line availability, can be as long as three weeks. Second, a limited number of documents are stored as full text, so, for the majority of documents, the only useful search criteria are author, title, date, and keyword. Moreover, there is uncertainty about what is and is not available for search. These latter two factors reduce search precision and recall, which are the basis of confidence in a system. What is required instead is on-line access, on-demand scanning, and full-text search immediately available to the technical staff.

2.3.2 Improvements to the Nuclear Document System

According to the NRC (Johnson and Moehle, 1993), other alternatives are open for accessing NUDOCS. It is possible to download these electronic documents on demand over AUTOS, but this process could easily lead to tying up communication access lines. A more reasonable approach would be to download HLW related documents loaded in NUDOCS with a batch process overnight. In order to establish this procedure, NUDOCS staff would need only to be informed of the requirements for batch downloading. While technically feasible, uploading of documents would become a possibility if legal issues related to electronic signatures were resolved. Support for this functionality would have to comply with NRC policies on electronic signatures.

In addition, there are short-term plans to make NUDOCS available on-line via Transmission Control Protocol/Internet Protocol (TCP/IP), perhaps sometime in 1994. There is also a long-term plan to revise NUDOCS. As soon as contracts are settled, AMS will begin a complete requirements analysis.

The NRC has recognized and worked toward overcoming inadequacies of current information management at the agency. IRM has contracted AMS to work with several agency offices to conduct a requirements analysis and to propose a document management system.

2.3.3 Document Management Requirements Analysis

On January 16, 1991, the Office of the Inspector General (OIG) issued a report recommending that IRM should use NUDOCS to meet the requirements of offices that maintain search and retrieval systems similar to NUDOCS. However, if those requirements could not be met, the OIG recommended that IRM propose feasible approaches to meeting them. The offices involved were the Office of the Secretary (SECY), the Atomic Safety and Licensing Board Panel (ASLBP), the Advisory Committee on Reactor Safeguards, and the Office of the General Counsel (OGC). IRM contracted AMS to initiate a requirements study to address these concerns (Chery, 1992).

The following is a summary of the work done by AMS as described in their requirements analysis report (AMS, 1991). In order to conduct the requirements analysis, AMS interviewed individual staff members in the offices noted above as well as others throughout the NRC who use NUDOCS. These users indicated the need for two types of support for their text management operations. The first area concerned specific features typical of any computer system used for this purpose. The second area concerned data structures required to adequately describe stored documents. The LSS needs analysis and conceptual design (completed by the Science Application International Corp. under contract to the U.S. Department of Energy, Office of Civilian Radioactive Waste Management, Contract DE-AC01-87RW00084) was also used as a framework. Features were divided into six categories: General, Query/Search, Retrieval, User Interface, Inputs, and Outputs.

The data structures required for document storage relate directly to the purpose of the individual systems. These purposes can be categorized into structured bibliographic record searching and full-text retrieval. The former assumes the user already knows something about the desired document, while the latter assumes the user knows very little and needs a very flexible approach. The two approaches require quite different underlying data structures. Specific features and structures are not presented here. The features and structures are, however, laid out in Table A-1 in the Appendix. The table is derived from the analysis conducted for IRM by AMS of the needs of various NRC Offices for office-controlled document management databases, and based on that study measures NUDOCS and a prototype system against these requirements. Together they are used to indicate features and data structures implemented by NUDOCS and those requested by the NRC offices for prototype proof of concept. This section will focus on findings of the present study.

In its report on requirements, AMS found that NUDOCS would require major enhancements and that the NRC would need to make a greater effort toward full-text capture to meet the requirements of its offices. The study suggested four alternatives:

- Expanded use of the current NUDOCS system
- Development of additional Oracle databases
- Redesign of NUDOCS
- Introduction of customer-controlled systems

In addition to the problems with NUDOCS pointed out earlier, the AMS study found the following problems in interviews with various office staff. Of the approximately 2 million records for documents submitted to NUDOCS, only approximately 35,000 documents have been scanned full text while only approximately another 35,000 include abstracts. In addition, it requires approximately one week to add an abstract to NUDOCS, and three to four weeks to add full text. Besides problems with timeliness, NUDOCS lacks the sophisticated capabilities of a full-text search and retrieval system. Some specific examples include its inability to route search output to the screen or a printer, perform proximity searches, rank hits and sort relevant headers, or provide a flexible and convenient user interface. It is not flexible enough to meet the needs of various users. These inadequacies are reported in full in the AMS requirements analysis report (AMS, 1991).

Examination of the four NUDOCS alternatives in terms of improvements and impacts favored the introduction of customer-controlled systems. These systems would be individually operated and controlled by each office, with data sharing among offices and NUDOCS. In order to investigate this alternative further, AMS recommended the design and development of a small prototype personal computer (PC)-based system.

In the prototype that followed, IRM elected to follow up on the fourth recommendation, combining the distributed capabilities of PCs with the minicomputer-based NUDOCS to provide customer-controlled functionality but to preserve the concept of NUDOCS as the central data repository (AMS, 1991; AMS and Pinkerton, 1992).

2.3.4 Prototype Document Management System

Following requirements analysis, under the direction of IRM, AMS designed and developed a prototype document management system (AMS and Pinkerton, 1992). A kickoff meeting reviewed the 67 features listed in the requirements study, 49 of which were rated as mandatory by at least one office; 37 of these were selected as being required for proof of concept. These requirements, along with assessments of NUDOCS and the prototype, are listed in Table A-1 in the Appendix as mentioned earlier (these tables combine a number of tables found in AMS reports). This section reviews the design, implementation, and results of the prototype system.

The prototype consisted of individual text management systems operated and controlled by the NRC Offices, which allowed sharing of relevant data. The prototype used full-text search and retrieval software for text management, installed for distributed processing on PCs connected on a LAN, and database software for data exchange with NUDOCS. The processing environment provided for the following:

- Local maintenance of bibliographic data, abstracts, and full text
- Sharing of this information with users throughout the NRC
- Submission of bibliographic information to NUDOCS

NUDOCS contained all records and full text of official documents, while local systems contained the documents local to each office and working copies of selected official documents.

The full-text software met many of the NRC's requirements, including an easy-to-use interface, proximity search, multiuser access, and update capabilities on the LAN. Furthermore, it allowed users to construct header records that could be passed in real time to NUDOCS and update the database, and to construct a search locally and then query and identify matching records in the database. It is important to note here that, while realtime updates to NUDOCS were shown to be technically feasible, the process required a good deal of custom code; also, nontechnical issues were avoided in this prototype by duplicating any tables that were accessed for update.

In addition to off-the-shelf software, a variety of custom software modules had to be implemented. This effort included parsers to translate queries between the full-text software and NUDOCS, Structured Query Language (SQL) generators, programs to receive query results and to download them as a file via SmarTerm, and procedures to search for documents in local databases. Approximately 400 programs contained in NUDOCS had to be examined to determine which were involved in query processing.

The text management prototype was evaluated from two perspectives:

- The technical ability to provide mandatory features
- The reaction of the NRC Offices in terms of the usefulness of those features

AMS stated that all 37 features selected for the prototype, and in fact most of the original 62 features, were successfully demonstrated. AMS and the NRC found that a customer-controlled microcomputer-based text management system with a communications interface to a central document repository is technically feasible. Nevertheless, it was also found that the employment of this concept, while technically feasible, does not provide a practical solution to agency-wide document needs. In fact, on review of the prototype, only the Advisory Committee on Reactor Safeguards requested that IRM provide a production system similar to the prototype at that time. However, the Advisory Committee on Nuclear Waste and Office of Nuclear Reactor Regulation (NRR) are presently implementing similar systems.

2.3.5 Proposed Document Management System

Following the above surveyed requirements analysis and prototype development, AMS has responded with a proposed document management system (AMS, 1993). The requirements proposed for this system are:

- Perform full-text search on and retrieval from their repository
- Access Certrec's RECALL generic regulatory documents
- Have multiple, simultaneous, local and remote access
- Interface to NUDOCS for search and retrieval in an integrated system
- Perform document imaging (i.e., scanning, OCR, and cleanup)
- Load documents from NUDOCS, imaging, and WordPerfect
- Have full functionality in the AUTOS environment

The Advisory Committee on Reactor Safeguards and the Advisory Committee on Nuclear Waste, in conjunction with a system at the National Institute of Health (NIH), currently use a DOS-based version of ZyIndex on Pcs for retrieving only reports they generate. The system provides for full-text search and retrieval of documents, but does not permit the incorporation of documents in reports without rekeying the information and does not provide access to generic regulatory documents or a central repository.

2.3.6 Office of Nuclear Reactor Regulation's Text Retrieval System

Somewhat similar to the AMS-proposed system is the Standard Review Plan - Update Project (SRP-UDP) Text Retrieval System now in use by the NRR. This system, marketed by Certrec, provides

ZyIndex software as the full-text indexing, search, and retrieval engine for a database of generic NRC documentation. The documents include:

- 10 CFR
- NRC Policy Statements
- Regulatory Guides 1-10
- Standard Review Plan
- NRC Circulars
- NRC Bulletins
- NRC Generic Letters
- NRC Information Notices
- LER Abstracts
- NUREG 0737

To meet their own needs, the NRR has had Certrec add:

- ABWR DFSER
- CE Sys 80+ DSER
- EPRI Evolutionary DSER
- EPRI Passive DSER
- Improved Standard Tech Specs (B&W, CE, GE4, GE6, WEST)
- NRC Inspection Manual (new with 9204 update)

Other differences with the system proposed by AMS are that there is no interface to NUDOCS through this system and no support for images. Nevertheless, the NRR system runs on AUTOS, its database is local, and it has standard access to NUDOCS via SmarTerm software.

Currently, there are 31 licensed users of the NRR system. The system is reportedly easy to learn and use. The only reported problem has to do with the fact that Certrec updates the database quarterly. On the one hand, the DHLWM would need updates more often than quarterly; on the other hand, Certrec does not support historical views on its updates.

While no requirements study was done prior to installing this system, it is interesting in that it supports the impression of a trend within the NRC (also noted in Johnson and Marshall, 1992). There is an obvious need for both structured queries and full-text search. Whether the interface is seamless, as in the case of the system proposed by AMS, or not, as in this case, NUDOCS is seen as adequate structurally for header queries, and ZyIndex/ZyImage is seen as adequate for full-text search. The standard approach seems to be a distributed processing one (as opposed to a distributed database approach), maintaining NUDOCS or its revision as the official repository while implementing in-house reference databases that (i) run under a graphical user interface (GUI), (ii) house unofficial or local documents, and (iii) come under the operation and control of each office, with selected sharing of data between offices (including the CNWRA). There is a need for enhancing NUDOCS, both its access and database, as noted in the next section. There also seems to be a common desire for access to the RECALL database.

2.4 DIVISION OF HIGH-LEVEL WASTE MANAGEMENT

According to the NWPA, the DOE is required to submit to the NRC a license application for the construction and operation of a mined geologic repository for the disposal of HLW. The DOE has collected data and information about the proposed repository site at Yucca Mountain since 1978. In order to fulfill its responsibility under the NWPA, the NRC has promulgated requirements in 10 CFR Part 60 necessitating both probabilistic and deterministic analyses. Thus, the DHLWM must prepare for a unique licensing situation. Following more than 10 years of prelicense application interaction with DOE, the NRC must review a repository license application in 3 years, allowing 18 months or less for technical evaluations and 18 months for licensing hearings (Johnson et al., 1992a).

2.4.1 The Advanced Computer Review System

In order to review the large amount of technical data accumulated, with much more expected, the DHLWM must implement appropriate computer capabilities and develop associated staff expertise. The DHLWM documented its functional needs for computer hardware and software to support its regulatory responsibilities and prelicensing consultation role with DOE (Chery, 1990). This system is called the DHLWM ACRS.

This section reviews work done by the CNWRA (Johnson et al., 1992a; Johnson et al., 1992b; Johnson and Marshall, 1992) analyzing requirements and proposing the ACRS for the DHLWM. A diagram of the ACRS and its related systems and networks is provided in Figure 2-3. The system is described in broad terms in order to show the place of TDOCS within it. Requirements for database access and document references are then reviewed.

2.4.2 Advanced Computer Review System Requirements

The basic requirements identified were technical computing capabilities and network connectivity. Networking and high-performance technical computing capabilities must meet DHLWM needs for a prelicensing technical review, licensing, and performance confirmation periods. Initiatives for the AUTOS (Gianios, 1991) and the high-performance technical review computer system in the DHLWM (NRC EDO, 1991) are complementary contributions to meeting these ends, but they introduce additional requirements.

ACRS requirements analysis, design, and implementation are concerned with five program activity areas. These areas are:

- Analysis method preparation
- Iterative performance assessment and review plans preparation
- Site characterization review
- NWPA regulatory requirements and guidance
- Management support

Analysis of needs to support these program activities identified five requirement categories for computer applications, upon which the design of computer hardware, network system, and specification of software are to be based. These requirements are:

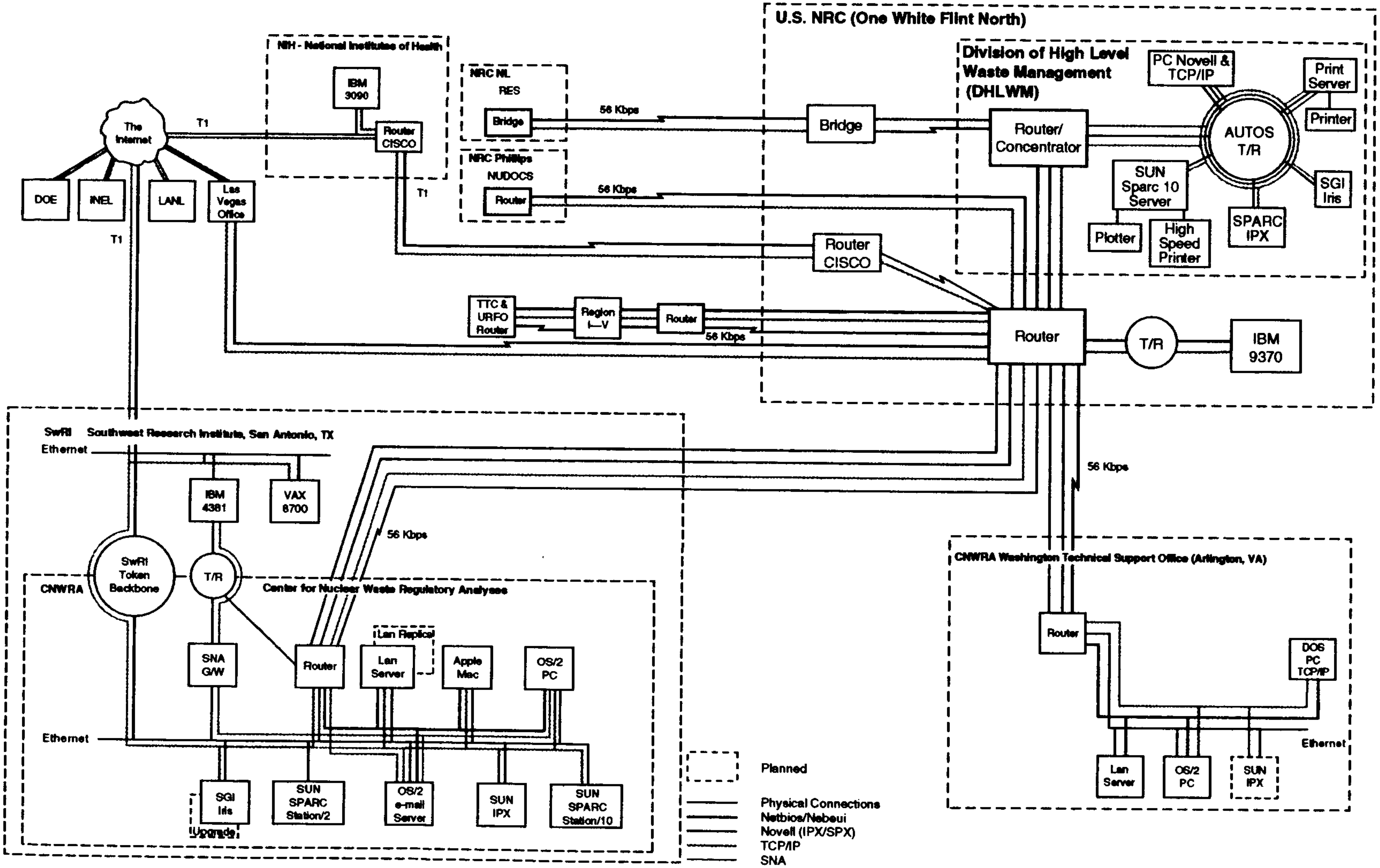


Figure 2-3. ACRS and related systems and networks

- High-performance technical computing
- Database access/document references
- Systematic regulatory analysis (SRA)
- Project management
- Office automation

2.4.3 Advanced Computer Review System and Document Management

This section identifies the place of TDOCS within ACRS. Its obvious place is in connection with database access and documents reference; however, document database management is also involved, directly and indirectly, in other requirement categories. Moreover, the CNWRA (Johnson and Marshall, 1992) recommended that any alternatives considered be coordinated with initiatives begun as a result of the January 16, 1991, report by the OIG. These initiatives by IRM are reported in Section 2.3.

The high-performance technical computing category implies several requirements related to TDOCS. This category, specifically related to analysis method preparation, deals with geologic and other program activities to analyze and display spatial and temporal data. For this purpose, the staff will need to display spatial data; construct feature models; compile and run a variety of engineering and scientific codes; and access large volumes of data from DOE, United States Geological Survey (USGS), and other databases and files. Part of this activity, related specifically to TDOCS, will be to electronically capture and format digital data including maps, charts, graphs, photographs, video, and other image data. This is also likely to be the case for iterative performance assessment evaluations. Likewise, rapid access to information relevant to the review and evaluation of DOE documents and other literature is required to support rulemaking and guidance activities.

Specific requirements related to database access and document references are many. For one, the NUDOCS system has been the interim system for loading, searching, and retrieving full-text HLW documents until the LSS is available. On-line access is required to make it effective for DHLWM and CNWRA use. The Technical Document Index (TDI) system in the CNWRA Library is used for control and retrieval of all hard copy documents. In it are indexed Regulatory Information Distribution System (RIDS) documents and technical reports produced by the CNWRA with reference documents for retrieval and review by the DHLWM. DHLWM staff will access and use databases containing waste package data citations and data reviews (over 1,000 citations with abstracts and about 110 data reviews formerly at the National Institute of Science and Technology (NIST)). Also, the DHLWM controls, manages, and provides staff with access to existing in-house technical reference databases, such as the Hydrologic Transport Section file, Site Characterization file, and others at the CNWRA and elsewhere. Furthermore, access to a number of industry and government databases is required. DIALOG and other on-line library services, as well as other databases internal and external to NRC need to be accessible on the agency's WAN. Finally, the DHLWM staff will use the LSS, when it becomes available, to provide document management and quality control for HLW documents, as promulgated in 10 CFR Part 2, Subpart J. Once it is operational, the submission of DHLWM documents to meet LSS requirements will begin. The volume of DHLWM external documents and image scanning will be reduced when the LSS is operational.

SRA, project management, and office automation entail additional requirements. PASS and PADB capture the thought processes whereby conclusions were reached and documented with references in the SRA process; thus, they are used as an investigative, analytical, and management tool and play a vital role in the licensing process. These applications require communications support between the

DHLWM and the CNWRA. In relation to project management, the CNWRA-developed Open Item Tracking System (OITS) and Commitment Control Log also require communications support between the DHLWM and the CNWRA. The need of the CNWRA and the DHLWM to access TDOCS further reinforces the requirement for communications support. Compatibility is an additional requirement implied not only by the need for communications between the DHLWM and the CNWRA but also by the need for TDOCS to interface with office automation requirements. AUTOS is designed to meet these final word processing, e-mail, and time management requirements.

2.4.4 In-house Databases and Other Technical References

The DHLWM also maintains three in-house document databases as well as other technical references. The three databases are repositories for Hydrology, Site Characterization Plan, and NIST/Materials documents. The technical references include position papers, journal articles, and other documents related to individual staff disciplines. Effective license application review will require control and management of these three databases and technical references, as well as access to them for incorporation in analyses and written reports. What follows is a brief description of these databases, along with their search and retrieval capabilities.

The Hydrology database is maintained by DHLWM staff in dBase III. Its index, in the form of bibliographic references and cross references to reviews, is updated as needed and kept on the office's file server, accessible to any staff member who has a copy of the dBase III software. The Site Characterization Plan database is a bibliography in WordPerfect format that can be freely distributed to staff. The NIST/Materials database is maintained by DHLWM staff in Advanced Revelation. This database consists of an index, abstracts, and approximately 1,000 reviews; it is maintained and made accessible to other staff on a dedicated PC. This system also contains 110 reviews of DOE documents, contained in 20 or more fields (and 3 files), that permit unique precise access to their contents of pertinent technical information and reviewer comments. A major use of these databases is to access references and incorporate them into analyses and to generate reports.

These three databases provide structured, keyword search capabilities on author, title, and other fields. Search results are bibliographic references and accession numbers used to locate paper documents kept on file in the office. Staff must key in the text in order to incorporate technical references into reviews. Full-text search and retrieval are not available for these documents.

2.4.5 Document Management Concerns and Capabilities

In discussions with DHLWM staff, a number of concerns with NUDOCS were expressed (Johnson and Moehle, 1993). For those DHLWM documents stored in NUDOCS, available search and retrieval capabilities are limited, and the interface is not user friendly. NUDOCS does not store electronic images. Dealing with hardcopy and microfiche, especially having to re-key document references into reports and reviews, is tedious and time consuming. The staff are not able to conduct research and review documents effectively following currently prescribed methods of document control.

These general capabilities are required by the DHLWM to perform its tasks:

- Consolidation of in-house databases, and scanning and loading those documents on demand
- Scanning and loading NRC paper documents and various other electronic data, and sharing data with the CNWRA
- Performance of both structured query and full-text search on technical references and incorporating them into reports and reviews
- Handling documents according to clearly defined policies and procedures with a user friendly interface

2.5 DOCUMENT MANAGEMENT AT THE CENTER FOR NUCLEAR WASTE REGULATORY ANALYSES

As addressed earlier, there is a need for the DHLWM to access the CNWRA's PASS/PADB system. This system, however, also serves as another model for document management systems, which, while its main purpose is document generation, is in many of its requirements quite relevant to the design of TDOCS. This section provides a brief overview of PASS/PADB and plans for its revision before examining in more detail its relevant document management requirements.

2.5.1 Program Architecture Support System/Program Architecture Database

PASS supports access to the PADB at the CNWRA. It was conceived as an integral part of the DHLWM Program Architecture (PA) and its support process of SRA. Its history depicts a dynamic system evolving to meet programmatic requirements and conceptual models (DeWispelare et al., 1992). It was implemented in 1988; Versions 1.0 and 2.0 were loaded with regulatory requirements, regulatory elements of proof, and regulatory and institutional uncertainty records, all derived from SRA. Version 1.0 supported the definition of initial regulatory requirements and the analysis of regulatory and institutional uncertainties. It was revised in 1990 to reflect changes in requirements and data relationships. Version 2.0 has become outdated because of further changes in requirements and data.

A proposal has been approved and implementation authorized for PASS/PADB Version 3.0 (DeWispelare et al., 1993). This proposed system design will permit storage and maintenance of PADB records and greater flexibility in the retrieval and formatting of PADB information in response to queries and reporting requirements. The design calls for using commercially available software — a full-text search and retrieval package, Verity's TOPIC; a SQL-compliant database, Oracle's Relational Database Management System (RDBMS); a GUI development tool, Visix's Galaxy; and a LAN/WAN-based client/server architecture.

These software packages were chosen because they support standards. TOPIC, Oracle, and Galaxy, for example, are all compatible with Microsoft Windows, SUN Open Look, IBM OS/2, and Apple Macintosh. They were also chosen because they support the notion of open systems, which is important when attempting to integrate various applications on various platforms in a networked client/server architecture. The TOPIC interface, for example, can be customized to provide for full-text search and structured query, external application launching, image file display, and transparent database access. Standards and open systems go a long way toward ensuring not only technical compatibility, but also the ability to economically modify the system as future requirements emerge.

2.5.2 Version 3.0 Requirements

Requirements analysis for PASS/PADB Version 3.0 takes a somewhat different approach from that of LSS and the AMS prototype. Specific features and structures will not be presented here. This section, instead, examines the requirement categories, and the manner in which they were related to system functions and constraints.

The requirements for PASS/PADB Version 3.0 fall under the following categories:

- Creation of textual materials
- Maintenance of textual materials
- Storage of textual materials
- Retrieval of textual materials
- Analysis and review of textual materials
- Reporting of textual materials
- Control of textual materials

The creation of technical materials requires word processing software, cut-and-paste facilities, network file server services, e-mail, and electronic file transfer. The maintenance of those material requires supporting unique material identifiers; inserting, modifying, and deleting materials; maintaining a history of changes, and accessing current versions while new versions are being prepared; and providing "check in/check out," and other configuration control procedures. Storage of those materials requires the selection, implementation, and identification of units as records so that they can be retrieved and reconstituted in a variety of sequences and groupings. Retrieval of those materials demands support for header-based, full-text, and concept-based searches. The requirements for analysis and review include a windowed GUI, scrolling, variable font and window sizes, and support for images (e.g., formulae, charts, and pictures). Reporting adds additional requirements for identifying and storing internal and external formatting information and parsing, identifying, storing, and retrieving low-level textual entities such as sections, paragraphs, or sentences. Requirements for control include access control, update control, and version tracking.

Many of the details underlying these requirements need to be considered in defining requirements for TDOCS. It should be noted, however, that there is emphasis in the PASS/PADB Version 3.0 requirements on textual materials and breaking them down into units for automatic report generation. TDOCS will be needed to support full-text search and retrieval of reference documents in the creation of SRA records. While it will be used to generate statistical reports and bibliographic listings, TDOCS will not, however, deal with automatic document generation, so most of the requirements concerned with storage and reporting can be ignored. Moreover, there will be a greater emphasis on images in TDOCS. Insights into requirements for managing image-based documents are presented in the section on LSS (Section 2.1).

2.6 SUMMARY OF RELATED SYSTEM CONSIDERATIONS

In attempting to survey the large number of document management systems within and outside of the DHLWM, many diverse requirements have been included. The unique and pressing requirements that justify TDOCS are:

- *Limited capabilities* to support the DHLWM technical review and ultimately compatibility and interface to the LSS
- Support for *header and full-text search*
- Support for *images* of non-textual materials
- Support for *on-demand scanning* to meet immediate needs
- Support for the *incorporation of materials and references* in analyses
- Support for *connectivity* with other systems and databases

Table 2-1 provides a summary of these related system policy and compatibility considerations for TDOCS. Since the status of and plans for these systems are changing continuously, these considerations are stated tentatively, and it is anticipated that further discussion will be necessary with specific parties responsible for each of the systems as design and implementation of TDOCS proceeds.

The question naturally arises, when considering TDOCS connectivity with NUDOCS and LSS/INFOStreams, as to whether TDOCS will have both upload and download capability. The answer to this question is somewhat complicated by implementation schedules and the official stature of HLW-related documents. The current NRC plan is not to upgrade but to revise NUDOCS, and that plan is now in the same requirements definition stage as TDOCS is. The current plan for the LSS is for the DOE to implement and operate it, with NRC oversight, in conjunction with INFOStreams, a plan that is still being discussed. Any interface at all between TDOCS and these systems depends on the nature and schedule of their implementations. An important matter for policy determination is whether an interface should be made between TDOCS and the current NUDOCS now or its revision later. Even after interfaces are in place, however, it is not expected that electronic uploads will be permitted. NUDOCS is the official repository of NRC documents, and the NRC needs to control what goes into it. The current policy holds that document submission must be made with hardcopy to the DCD because NUDOCS is microfiche-based; any changes to that policy would need to be determined by the NRC. Document downloads are much more likely to be possible, but whether these can be made on-demand or in batch overnight is another matter to be determined. Much the same can be expected for the LSS when it becomes the official repository and HLW documents are fed from NUDOCS to it. Even though it is not expected that uploads will be permitted, uploads and downloads, and the precise nature of such mechanisms, will be a matter of NRC and possibly even DOE policy.

Another related question arises when considering connectivity between TDOCS and the systems discussed in this section, and that is whether documents should be actually downloaded and incorporated into the TDOCS database or referenced by means of dial-in access to those systems. This is another matter for policy decision, but the following arguments seem to favor the download and incorporate approach. NUDOCS is not now and it is not certain whether the NUDOCS revision or the LSS/INFOStreams implementation will be set up for download or dial-in access to the full text and images of documents; it is more likely that full documents will have to be transferred by means of microfiche or electronic media such as tape or disk. If documents are not incorporated into the TDOCS database, then it is impossible to embed hypertext or notes in them or to perform full-text search and retrieval on them. The consequences of incorporating all relevant documents are, however, the costs of storing large volumes of data and the redundancy of storing documents in two systems, one official and the other not. However, it is envisioned that TDOCS will be limited in size due to its emphasis on supporting technical review prior to the availability of the LSS. Only about 50 percent of all documents related to the HLW program will be technical in nature. Of these technical documents, only strategically selected types will be loaded into TDOCS. Cost and resource constraints will also limit the size of TDOCS.

Most of the considerations in Table 2-1 have been incorporated into the requirements summary table at the end of Section 3. Specifically, the last column headed "Dependent on Schedule or Facilities of Other Systems" addresses these as a function of the broader set of overall requirements for TDOCS. It will be important to discuss the following considerations with the responsible parties and make certain that policy and procedural implications for TDOCS for these related systems are clearly understood and agreed upon by the DHLWM, IRM, and the CNWRA.

Table 2-1. Related system policy and compatibility considerations

Related System	Policy and Compatibility Considerations for TDOCS
LSS	• Control of DHLWM hard copy, ASCII, and image submission to the LSS • Technical data package header/table of contents capability • Header and full-text search capability • Hypertext links to images • Schedule for LSS implementation monitored for later TDOCS expansion • Access and retrieval plan for LSS in DHLWM
DOE/INFOStreams	• Rule-based system for inclusion of DOE and other parties' documents for LSS • Form, period of value, and archival considerations for licensing documents • LSS documents captured by DOE due to large volume (85 percent of its own and 15 percent for other parties) • QA of documents essential for licensing • Audit trail, pedigree of information, and defensibility of what was collected and how parameters are laid over the technology used for processing • License application developed as one stream with pointers to supporting data • Transfer of documents required by the DHLWM to TDOCS • Categories of information at 4 levels: primary, supporting, processed, and preliminary (raw data) • Activities producing information in HLW program – site characterization plan – licensing application – general – safety analysis report – environmental impact statement – QA, Health Physics, and Emergency Plans
DOE/Records Management	• Maintenance of information in current state rather than archived • Understanding the licensing process and the information imposed by that process • Capturing and reviewing huge volume of information in 3-year time frame • Electronic transmission of filing, order, and decisions

Related System	Policy and Compatibility Considerations for TDOCS
NRC/NUDOCS	• Upload (signature problem)/download (overnight) capability to TDOCS • Document submission procedures compatibility • Archival requirements for DHLWM where different • NUDOCS upgrade to full text and image system impact on TDOCS • Official repository for agency documents • NUDOCS upload/download requirements (from TDOCS direct to LSS or through NUDOCS to LSS)
NRC/Office Databases	• Structured, bibliographic header queries and full-text searches • Staff-controlled systems • Distributed database on client server • Local maintenance of headers, abstracts, and text • On-line query to NUDOCS for matching records • Custom software for interface with NUDOCS • Not needed by all offices • Load documents from NUDOCS • Have full functionality in AUTOS environment
DHLWM/ACRS	• Implement on the existing Sun Sparc 10 server or similar computer • Electronically capture and format digital data from printed graphic and text materials, including maps, charts, graphs, photographs, video, and other image data • Rapid access to and importing of text from reference information and other file information relevant to the review and evaluation of DOE documents and other literature required to support rulemaking and guidance • On-line access to NUDOCS • Waste-package data citations and data review in NIST • In-house technical reference database, such as the Hydrologic Transport Section file, Site Characterization file, and others at the CNWRA and elsewhere • Access to a number of industry and government databases required • Dialog and other on-line library services • Interface to PASS records as investigative and analytical management tool • Interface to the OITS and Commitment Control Log • Communications and office automation support

Related System	Policy and Compatibility Considerations for TDOCS
PASS/PADB	• Transfer of documents on-line • Interface to PASS for creation, maintenance, storage, retrieval, analysis and review, reporting, and control of textual materials • Compatible word processing interface, cut and paste, network e-mail, and electronic file transfer • Windowed GUI interface

3 REQUIREMENTS ANALYSIS

This section defines requirements for TDOCS. Based on this and previous analyses of the DHLWM and its technical staff's current and evolving needs to load, access, use, and manage documents in an in-house technical reference document database. These requirements also draw on efforts to define requirements, design, and implement other document management systems on the part of the DOE, the NRC, and the CNWRA, including LSS, INFOStreams, Improved Records Information System (IRIS), NUDOCS, IRM Prototype, PASS/PADB, and TDI.

The strategy is to provide a system capability that meets stated requirements and is currently achievable and immediately beneficial. What is currently achievable is best approached through the deployment of the latest document database technology. This is best accomplished for this task with off-the-shelf software packages that meet industry standards where possible, custom software where not. What is immediately beneficial, in a technical reference document database, is a system that ensures confidence and value. Confidence, as defined earlier, has to do with whether the system has been loaded with and provides access to all and only those document references relevant to the user's specific needs. Value has largely to do with whether the system, once appropriate document references have been found, provides all the functionality necessary for users to perform their day-to-day tasks. Moreover, an achievable and beneficial system must be accompanied by policies, procedures, and responsibilities that clearly define standards for loading, accessing, using, and managing the system. Only by enforcing system policies, procedures, and responsibilities can there be any assurance of continued confidence and value as the system evolves to meet future needs.

This section identifies requirements in terms of functions, constraints, and policies that are applicable to the TDOCS system. There are five major functions:

- Database loading: getting documents into the database
- Document processing: cleaning up, entering headers, and indexing full text
- Search and retrieval: accessing the documents
- Document manipulation: facilitating document use
- Administration and maintenance: assuring system functionality

These functions should be implemented with a GUI, on multiple platforms, in a client/server architecture using commercially available software packages where possible. Implementation under these kinds of constraints aims at minimizing the impact on existing and planned systems and configurations, maximizing expandability to meet evolving needs, and conforming to NRC software development policies. It is important to point out that these constraints are imposed by the AUTOS and ACRS systems in use by the DHLWM, by the CNWRA's PASS/PADB and OITS system interfaces for DHLWM staff, and by NRC policies. As has been stated previously, the strategy is to provide a system capability that is currently achievable and immediately beneficial. Efforts to make TDOCS interfaces completely seamless, provide for easy access to all document databases, and support imaging functions should be based on, among other things, identifiable technical alternatives. For the most part, these constraints do not minimize but rather maximize the potential benefits of TDOCS for the DHLWM staff.

At this point, many TDOCS functions are open-ended. Specific details of functionality simply cannot be defined precisely until off-the-shelf software packages have been selected during design. Many functions must be defined in terms of alternative approaches and actions that can be decided on as policies, procedures, and responsibilities that are determined cooperatively between the DHLWM and the CNWRA during system design. It is these policies, procedures, and responsibilities that, again, ensure the quality of the documents and confidence in the system. Thus, this definition of requirements delimits what functionality TDOCS should be expected to support, how that functionality should be constrained both in design and implementation, and what policies, procedures, and responsibilities should be enforced in use and management.

3.1 SYSTEM FUNCTIONS

3.1.1 Database Loading

Selected documents, rather than the entire backlog, should be added to TDOCS by means of routine loading, on-demand loading, and electronic downloading of technical references. Approximately 15,000 documents have been accumulated in the CNWRA library over the last five years. Based on experience at the CNWRA and considering resource constraints and activity schedule of the DHLWM, it is estimated that the DHLWM might be able to add approximately 20 documents per day. This estimate includes scanning, OCR, cleanup, header entry, and full-text indexing. At this rate, approximately 10,000 documents could be accumulated into TDOCS over a period of two years. It is further estimated that each document would consist of on average 100 pages, and that 10 kilobytes would be required for storage of each page. This estimate includes storage of text, images, header, and full-text index. Thus, approximately 10 gigabytes of storage space would be required for the estimated 10,000 documents, which would have little more impact on the existing ACRS than perhaps requiring an additional external harddrive.

Section 3.1.1 will focus on the sources of system data and introduce some of the costs and tradeoffs involved in loading the documents. Section 3.1.2 will discuss a number of document processing functions that support database loading, such as scanning, OCR, and cleanup, image capture, header entry, and full-text indexing. Constraints on loading and processing as well as policies, procedures, and responsibilities specifying the scope of documents to be loaded, quality of scanning, OCR, and cleanup; the degree to which document formatting and images can be preserved; and the need for bibliographic headers will be discussed in later sections.

3.1.1.1 Routine Loading

Routine loading should be implemented through the CNWRA, contractors, and/or the DHLWM secretarial staff in order to support document scanning, OCR, and loading functions on a scheduled basis. This should be the primary method of entering materials into the text and image repositories. Requests for routine loading should originate in several ways:

- Certain materials, such as reports that the DHLWM receives from the DOE, should be submitted for routine scanning, OCR and loading

- Certain document sources, such as particularly relevant technical journals, may be identified for loading; articles from those sources should be identified with the staff member who requested that it be loaded and submitted for routine scanning, OCR, and loading as they are received

- Staff may identify additional materials, such as technical reports, that need to be captured and entered into the system; these materials should be submitted for routine loading even though selected pages of such materials may have already been captured through on-demand loading

3.1.1.2 On-Demand Loading

On-demand loading should be implemented to support immediate staff requirements for limited amounts of textual and/or graphical materials. A typical request for on-demand loading would arise when a staff member needed several pages from a report or journal article to incorporate into another document. The required pages could be scanned by the technical staff member or given to a secretary for on-demand loading. In this way, staff would be able to load small numbers of pages of relevant materials, identified with the staff member who requested that it be loaded, so that these materials could become immediately available on the system.

As soon as the on-demand scanning process is completed, the scanned material should be made available to the staff. However, the materials scanned on an on-demand basis should not be loaded into the permanent repository, because they would generally represent partial or incomplete documents. Therefore, on-demand loading should always be coupled with a concurrent request for routine loading of the full document from which the on-demand pages were selected.

3.1.1.3 Electronic Loading

It is anticipated that some of the materials to be loaded in the system will be available in electronic form. This is particularly true for materials generated internally by the DHLWM or CNWRA. Such materials are anticipated to be available in electronic form as full-text with accompanying bit-mapped files of figures, equations, images, etc.

Tape, Optical Disk, etc. Materials that have been created in appropriate electronic formats may be entered directly, bypassing the scanning and OCR processes entirely. These materials may be received on a variety of electronic media including magnetic tape, diskettes, optical disks, or via communications facilities.

Other Databases (NUDOCS, PASS/PADB, IRIS). Documents and materials required by TDOCS users may be available as electronic copies from other databases and systems that support full-text and images. However, careful consideration must be given to the handling of graphical images. An external database or system may contain a number of relevant documents in full-text form, but if the associated images are not present in a compatible format, electronic entry of these materials into TDOCS may not be possible in their entirety. That is, if an image format incompatibility exists, then only text would be downloaded. Access to NUDOCS, PASS/PADB, and other databases should be automated and made available to staff through TDOCS menus. Access to other databases, for example IRIS and related technical databases, will be determined later. Implementation should be flexible and expandable to facilitate staff in gaining access to additional databases.

On-Line Bibliographic Services. A number of bibliographic services are available that provide additional support for staff research and document preparation. Many of these bibliographic services provide a download capability by which the headers, and occasionally the full text, may be retrieved in electronic form. Access to at least two of the most commonly used bibliographic services (to be identified at a later date) should be automated and made available to staff through TDOCS menus, using the communications package supplied with AUTOS. The implementation of this facility should be flexible and expandable to facilitate staff in gaining access to more bibliographic services in the future. Interfaces should be provided for capturing headers and full-text from bibliographical services. However, the potential limitations, mentioned above, associated with the absence or incompatible formatting of graphic images, may make it infeasible to download the document in its entirety. That is, if an image format incompatibility exists, then only the text would be downloaded.

Where possible, bibliographic headers and references should be obtained electronically from other systems to avoid the required effort and potential for error associated with creation of new bibliographic headers. In practice, the ability of TDOCS to capture and download headers and references may be limited by incompatibilities in the header formats between systems and/or limitations in the system interfaces. It is expected that some custom code would have to be implemented to handle the interpretation and reformatting of headers between the bibliographic services and TDOCS.

The DHLWM will also need to acquire selected DOE technical data packages, or at least the documents contained within them, before they are available through the LSS/INFOStreams. DOE is currently writing headers and tables of contents for existing packages. Headers are presumably stored in a database and provide cross-references to microfilmed material and storage locations for non-scannable materials. For meaningful access, however, the tables of contents would also be required from the records management system. TDOCS should, thus, provide download access to available technical document package headers and tables of contents when these are available through LSS/INFOStreams.

3.1.2 Document Processing

Routine, on-demand, and electronic loading of technical references will require document processing. This section will focus on a number of document processing functions that support database loading, such as scanning, OCR, cleanup, image capture, header entry, and full-text indexing. Constraints on processing, as well as policies, procedures, and responsibilities concerning the quality of scanning, OCR, and cleanup; the degree to which document formatting and images can be preserved; and the need for bibliographic headers, will be discussed in later sections.

3.1.2.1 Scanning, Optical Character Recognition, and Cleanup

The process of scanning materials and converting them to full text through OCR is subject to errors that may arise from several conditions. Scanning and OCR errors may or may not be detected automatically, and a manual document cleanup procedure is normally required as part of the document capture process.

Generation Loss. Generation loss occurs as a result of documents that have lost resolution or quality as a result of being reproduced one or more times. This commonly occurs when copies of other copies are submitted for scanning. The edges of characters may be broken, distorted, or merged with adjacent characters, causing the OCR process to fail. Similarly, graphical images, and particularly photographic images, may be highly distorted by multiple generations of reproduction. Therefore, it is important to ensure that first-generation materials, such as original copies of printed reports, are used wherever possible for system input.

OCR Effectiveness. The effectiveness of the OCR process is affected by both the quality and content of the input document. Scanned documents that have broken or touching characters may cause the OCR process to misidentify, or completely fail to recognize one or more characters. While OCR with good quality input documents is usually more than 99 percent effective, that level of reliability implies an average of up to thirty character errors per page. Additional factors that may adversely affect the OCR process include intermingling of text and graphics, text that is oriented at angles, handwritten notes or marginalia, etc. Therefore, it is important that the OCR process be as effective as reasonably achievable and that it be coupled with an appropriate document cleanup process to ensure user confidence in the fidelity of the converted textual materials.

Preservation of Formatting and Images. Graphical images, figures, tables, and equations must be handled appropriately during the scanning and OCR processes. These non-text objects must be recognized, identified, stored in an appropriate format, and properly associated with the textual materials. Selection, identification, and processing of graphical images is normally accomplished interactively by an operator during the document processing stage. In addition to graphical objects, certain formatting, such as superscripts, subscripts, italics, bolding, underlining, etc., may appear in the text of the source document. These formatting attributes are normally lost when the OCR process converts a scanned image to ASCII text. Therefore, a word processing format rather than ASCII text should be considered for the output of the OCR process so that such formatting attributes may be retained.

The preservation of document formatting and images is a major requirement for TDOCS. Selected software packages, such as the full-text indexing, search and retrieval engine, must either support this directly by maintaining formatting and images or indirectly by supporting hyperlinks embedded in text so that relevant and essential formulae, equations, graphs, charts, pictures, and other images can be stored in separate files, accessed, and displayed in windows alongside the text.

Cleanup Accuracy. Since the cleanup process involves human intervention, a question may arise about the accuracy and fidelity of the resulting scanned materials. For example, when the OCR process fails because of distorted characters in the input document, the cleanup process will require an operator to correct the error. If the operator is not familiar with the terminology or subject matter, the wrong word can easily be substituted where the OCR or scanning failed. Thus, the training and experience level of the person providing the cleanup function can have a significant effect on the accuracy of the text derived from a scanned document. One approach to this problem is to simply flag and count the errors and indicate them to the user. This effectively bypasses the cleanup process and transfers the responsibility for interpretation of the scanned materials to the user. Even though this approach may not be the most acceptable, it may be necessary due to resource constraints. The TDOCS system must address the question of cleanup accuracy in a way that is unambiguous to the user by flagging documents that have not been cleaned up and by assuring that cleanup, when performed, is complete and accurate.

3.1.2.2 Bibliographic Header Entry

Bibliographic headers should be required for all TDOCS entries to permit searching for documents using author, title, journal name, volume, issue, date, and other standard fields. The intent of the bibliographic header is to enable the user to quickly determine whether a selected document is likely to contain relevant information.

Textual Documents. It is expected that most TDOCS documents will be primarily textual in nature, containing a limited number of images and figures. Normal bibliographic headers, including such information as title, author, journal name, volume, issue, report number, date, publisher, etc., should be prepared for these materials. When available, the abstract for textual documents should also be entered.

Non-OCR Images. It is expected that some TDOCS documents will consist of images that cannot meaningfully be converted to text. Examples of this type of document would be a geologic map, a photograph, or a frame captured from video tape. Bibliographic headers including a title or description, date, size, and other attributes, may need to be prepared for these documents when they are incorporated into technical reviews and reports. This does not imply a need to maintain bibliographic headers for images contained in the DHLWM technical database, though consideration might be given to providing cross references between databases. This will be a matter for further discussion between the DHLWM and the CNWRA.

Non-Scannable Material. It is expected that some materials, such as data on magnetic media, that are not scannable will be referenced in TDOCS. Bibliographic headers for these materials will provide the only method for identifying and describing them. Therefore, headers for non-scannable materials should be more descriptive than other headers, containing such fields as description, date, media type, recording parameters, etc. As with non-OCR images, bibliographic headers may need to be prepared for these documents when they are incorporated into technical reviews and reports. This does not imply a need to maintain bibliographic headers for images contained in the DHLWM technical database, though consideration might be given to providing cross references between databases. This will be a matter for further discussion between the DHLWM and the CNWRA.

Accuracy and Completeness. Because the header entry may be performed by the operator during the scanning process rather than by a trained cataloger, the header formats should be designed to facilitate entry from information that is readily available on the document. The accuracy and completeness of the header should be verified by a visual review by the operator at the completion of the header entry process.

Bibliographic headers should be entered by the operator as part of the document loading process. A specialized entry screen, specific to the type of document being entered, should be displayed, and the operator should be able to "fill in the blanks" to complete the header. The content of bibliographic headers should be limited to information readily available on the cover or first few pages of a document so that specialized cataloging skills will not be required for entry of header information. Thus, header fields that require specific subject knowledge to complete, such as "keywords" or "subject," might be optional in the bibliographic headers. Optional fields would be filled in if such information is available, either found in the document itself, as is the case with many journal papers, or provided by staff who request the loading of a specific document and are more likely to know its contents. Decision on header content will be a matter for further discussion between the DHLWM and the CNWRA.

3.1.2.3 Full-Text Indexing

It is expected that, because the content of the bibliographic headers will be relatively constrained by the approach of preparing them from information readily available on cover or first few pages of a document, the users of TDOCS will rely heavily on full-text searches to find desired materials. Thus, full-text indexing of TDOCS materials is a particularly important function of the system.

Textual Documents. It is expected that most TDOCS documents will be primarily textual in nature and will be converted to full-text through an OCR process. Graphic images, including figures, illustrations, photographs, formulae, equations, etc., should be converted to bitmapped images that can be associated with the full-text documents for display purposes. The resulting full-text and its associated bitmapped image files should be loaded into a full-text repository in a format that permits search, retrieval, and display by a full-text processing system. Indexing of the full-text should be accomplished automatically as part of the loading process. Header information should be formatted into full-text data fields so that the documents may be retrieved through text searches of the header information.

Non-OCR Images and Non-Scannable Material. Materials consisting solely of images and materials, such as magnetic tapes or other machine readable media that are not subject to OCR or scanning, should be identified and described by bibliographic headers. These headers should be processed, formatted, and loaded into the full-text repository so that header information about non-OCR images and non-scannable materials can be searched and retrieved through the full-text system.

Although non-OCR images and non-scannable materials cannot be captured in the full-text system, substitute documents or slip-sheets should be prepared, including a brief description of the materials. These substitute documents would serve two purposes: (i) they would provide a visible and concise description of the materials, and (ii) they would provide a vehicle for the icon or hypertext link used to select the associated non-OCR images for graphical display. Thus, when a user searches for a particular geologic map in the full-text system, an entry would appear for the document in the full-text search results. Viewing that entry would cause the substitute document, including a concise description of the map, to be displayed. The user may then select the display icon or hypertext link on the substitute document to cause a graphic image of the map to display. However, if the user determines from the substitute document that the referenced map is not the one desired, he or she may proceed by examining other entries in the full-text selection list, and the time required for the graphic display of the map is avoided.

As stated in Section 3.1.2.2, bibliographic headers may need to be prepared for these documents only when they are incorporated into technical reviews and reports. This does not imply a need to maintain bibliographic headers for images contained in the DHLWM technical database, though consideration might be given to providing cross references between databases. This will be a matter for further discussion between the DHLWM and the CNWRA and possible prototyping.

3.1.3 Search and Retrieval

Once relevant documents have been loaded and processed, the staff must be able to find desired material quickly and reliably. A broad range of functions are needed to facilitate search and retrieval of materials stored in TDOCS. The degree to which TDOCS supports flexible, powerful, and reliable document access facilities will contribute significantly to the building of user confidence and, consequently, to increased staff productivity. Constraints on search and retrieval will be discussed in a later section.

3.1.3.1 Document Access

In order to ensure confidence in retrieval, TDOCS should support three types of document access: full-text search, structured header queries, and hyperlinks between documents.

Full-Text Search. Full-text search is the capability to search the text of an entire document for words, phrases, and combinations of words. Because of the limited content of bibliographic headers, full-text search ensures greater confidence in finding all relevant documents when a user performs a subject-oriented query. However, unlike structured queries, full-text search is more likely to find irrelevant material, because the search is based on specific words or phrases that may occur in many documents. The degree of confidence in full-text search depends only in part on the accuracy of search criteria. It also depends on the degree of completeness of the document database and the accuracy achieved in document scanning, OCR, and cleanup. Similarly, the full-text search may not detect occurrences of the specified search terms that are incorrectly represented in the database due to errors that occurred in the scanning, OCR, and cleanup processes.

Structured Header Queries. Structured header queries provide the capability to locate documents by specific attributes, and are particularly effective when the user is searching for a specific document or set of documents about which some information is known. For example, structured header queries can be used to find all correspondence written by a particular author, memos that refer to a specific topic, or journal articles with certain keywords in the title. An important advantage of querying headers is that headers can be created for documentary materials, such as images, data, microfilm, photographs, videos, data on magnetic media, etc., that are not textual and cannot be scanned. The degree of user confidence achieved by structured header queries depends only in part on the accuracy of the search criteria. It also depends on the quality, completeness, and accuracy of the headers entered into the database when documents are loaded.

Hyperlinks. Hyperlinks between documents provide an added degree of confidence in finding appropriate materials by allowing users to establish electronic relationships between documents. This will be especially useful to staff as they become familiar with the set of documents loaded into the system. TDOCS should provide the capability for staff to embed hyperlinks in a particular document that point to other documents. When the document is retrieved and displayed, the hyperlinks would be highlighted. By selecting a hyperlink in one document, the associated document would automatically be retrieved and displayed. This permits a user to identify and associate a number of documents containing related material so that they may be retrieved and traversed in an efficient and intuitive manner. Hyperlinks have an added advantage in that they permit the relationships between documents, identified by one staff member, to be captured and made available to other members of the staff.

Another important use of hyperlinks is the creation of tables of contents for sets of documents, such as LSS packages, or for documents, such as 10 CFR Part 60. These hyperlinks would maintain the organization of document sets or sections and allow staff to start with a table of contents, select and jump immediately to particular documents, document sections, or other materials. Where it is deemed useful for pre-loaded sets of documents, TDOCS should support the creation of readymade hyperlinks.

3.1.3.2 Search Confidence

User confidence is an important factor in determining the success of both document databases and search and retrieval systems. Just as document loading must strive to support completeness and accuracy of data, retrieval of that data must strive for precision and recall. The system must ensure, as much as possible, that all documents relevant to the task at hand, and only those documents, are found. User confidence is also enhanced by systems that sort the search results in an order that reflects the likelihood that the selected documents are relevant to the search criteria. In order to achieve a high level of user confidence, TDOCS should provide a variety of query and search techniques. The precise combination of search techniques provided in TDOCS depends to a large extent on which off-the-shelf software packages are selected, and particularly which software is selected for full-text search. However, the following techniques should be supported:

- Wildcards and Boolean operators: the capability to combine search criteria with logical operators such as AND, OR, and NOT to create more specific queries
- Near spell search: the capability to find words regardless of syntactic variety (number, tense, and other affixes)
- Fuzzy search: the capability to find words even though they may be misspelled in search criteria or in the documents themselves
- Phrase search: the capability to find not just words but phrases
- Proximity search: the capability to find combinations of words and to specify the distance between those words in a document (e.g., phrase, sentence, paragraph, etc.)
- Cross-partition search: the capability to search for documents of specified type or category
- Search result ranking: the capability to rank search results in display lists

3.1.3.3 Concept Search and Building

One of the major limitations of full-text search systems is that the user must be able to anticipate specific words used in the text. In many cases, this does not present a problem. However, when dealing with complex or technical subject matter, a wide variety of terms may be closely related to the subject of interest. Quite often, what one is looking for is not a word but an idea that requires several words in a phrase or within close proximity to express. Concept-based searches provide a means of defining the terms associated with a particular concept in a way that causes the search facility to automatically find documents containing any of the associated terms. A concept-based search can be constructed from various combinations of wild card and Boolean, near spell, fuzzy, phrase, and proximity operators. For example:

- A paragraph containing "operations" and ("environment," "closure," or "container")
- A phrase containing "60.111(b)"
- A paragraph containing "retrieve" and ("waste," "container," "canister," or "HLW")
- A paragraph containing "underground facilities" and ("fracture," "opening," or "stability")
- A paragraph containing "emplacement" and ("operations" or "container")
- A paragraph containing "backfill" and ("EBS," "Engineered Barrier System," or "Waste Package")
- A paragraph containing "thermal load" and ("stability" or "mechanical")
- A paragraph containing ("EBS" or "Engineered Barrier System") and ("Canister," "Container," or "Waste Package")
- A paragraph containing "recovery" and ("spent fuel," "value," or "resource")

The use of "concepts" permits users to accumulate and share knowledge about how best to retrieve information on specific subjects. In a real sense, the users, through their accumulated experience and the concept searches that they build, teach the system how to make efficient retrievals. Thus, all users benefit from the experience of other users through the development and retention of concept-based searches.

In short, concept building is an important means of locating documents. Concept building means the staff would have the ability to build their own concepts for use in document search. This capability would become more and more important as the staff become familiar with the sets of documents contained in TDOCS and with the various idioms and technical expressions different professionals employ for similar concepts.

3.1.3.4 Query Save, Recall, and Edit

Another means of supporting user confidence is through the saving, recalling, and editing of search and query criteria. Search and query are seen as iterative processes. Staff should be able to progressively refine (i.e., narrow and broaden) criteria until they are satisfied that the resultant set of documents represents what they are looking for. Staff should also be able to save search and query criteria so that they can be recalled and reused by themselves or other staff members.

3.1.3.5 Search Result Browsing

Search result browsing is another means of supporting user confidence. The results of a structured header query or a full-text search should be displayed as a list of found documents so staff can select documents from that list for viewing. The text of the selected document would then be displayed in a separate window, which can be scrolled, line-by-line or page-by-page to view portions of the text not currently visible in the window. The graphical user interface should provide "slider-bars" to permit the user to move rapidly to any portion of the text of the displayed document. Text search facilities should also be provided to enable the user to find additional specific words in the displayed document. Facilities should also be provided to enable the user to display successive documents without having to return to the selection list.

3.1.3.6 Concurrent, Multiple Document Viewing and Scrolling

Concurrent, multiple document viewing would be useful for comparing multiple documents or extracting material from multiple documents to support analyses. Staff would be able to select multiple documents from a search result list and view them at the same time in separate windows, which may be individually selected, scrolled, or searched to identify the desired text.

3.1.3.7 In-Document Match Highlighting and Browsing

In-document match highlighting and search capabilities can also support user confidence. When a document is selected for viewing from a search result list, the search terms matched in the document text should be highlighted. Facilities should also be provided to permit the user to move rapidly, forward or backward, between successive highlighted search terms. This would permit staff to quickly find the portion of the displayed text that pertains to the query, and then to move through the document identifying any successive matches.

3.1.3.8 Hyperlink Creation

Hyperlinks between documents has already been discussed as an important means of locating documents. Providing for hyperlink creation would mean the staff could create their own links between documents. This capability would potentially become more and more important as staff become familiar with the documents contained in TDOCS and begin to see associations among them.

3.1.4 Document Manipulation

A broad range of functions are needed to make TDOCS useful to the DHLWM staff. Once relevant documents have been found, there are a number of functions that need to be performed. Many of these functions involve manipulating the documents. Some are essential to the work of the staff, such as cutting and pasting to incorporate documents into analyses, and some are simply useful, such as embedding notes into documents so staff can continue their typical research habits in an automated environment. Other functions included under this broad heading would enable staff to enrich the confidence of locating documents, such as creating hyperlinks and building concepts. Constraints and policies concerning these functions are discussed in later sections.

3.1.4.1 Cut, Copy, and Paste

Once a relevant document has been found and displayed in one interface window, staff will need to be able to cut or copy portions of it to an interface clipboard, and from there paste it into another document in a word processor. This is an important function in that it should satisfy the staff's need to be able to incorporate documents into technical analysis.

3.1.4.2 Document Printing

While it is the purpose of TDOCS to eliminate paper in the DHLWM, it is also recognized that document printing will be required. TDOCS should provide support for the printing of document text directly from the full-text display facilities.

3.1.4.3 Report Generation

Report generation is currently a part of the Hydrology, NIST/Materials, and Site Characterization databases. This capability should be incorporated in the TDOCS database facilities. This functionality provides users with control of bibliographic data and permit staff to search for bibliographic references and extract bibliographic entries in proper format for literature reviews.

Other types of reports are also possible. These reports might include bibliographical listings of documents of a particular type, documents written by a particular author, documents written within a range of dates, and so on. One type of report that might be particularly useful would be a listing of newly loaded documents; this sort of report could be generated, say, weekly and sent via e-mail to the entire staff.

3.1.4.4 Document Download and Transfer

The DHLWM staff will need to download documents to their own workstations and send documents to one another over the LAN via e-mail attachments. These capabilities are provided by the currently existing AUTOS and ACRS systems but will require some custom code to automate.

3.1.4.5 Public and Private Notes

TDOCS should support the embedding of notes within documents. Notes would allow staff to mark documents with marginalia so that any time the document was viewed they could also view margin notes. It should be possible to set this function up so that staff can create both public and private notes. Public notes are those that are shared among staff; that is, anyone can view them. This could present a problem if everyone enters notes into documents since notes are typically meaningful only to the person who enters them. Thus private notes, those that only the one who enters them can view, can be used. The ability to make notes public and private would depend on the ability to set up user accounts.

3.1.4.6 Image Manipulation and Enhancement

The DHLWM staff will need at least a limited amount of image manipulation and enhancement capabilities in order to allow them to view a complete image and "zoom in" to view a selected portion of the image in greater detail. Graphic displays, particularly of maps and photographs, require high-resolution display hardware for viewing.

3.1.5 Administration and Maintenance

For the purpose of administering and maintaining the system and measuring its impact on existing and planned systems and configurations, TDOCS should implement a standard set of tools and reports. In addition, a set of policies, procedures, and responsibilities should be specified as guidelines for system administration and maintenance. These policies are discussed further in a later section.

3.1.5.1 User Passwords, Privileges, and Accounts

For the purpose of administering and maintaining TDOCS, a number of tools should be implemented. These tools should provide for system password access, document and database modification privileges, document tracking and configuration control, and backup and recovery. In some cases, the need for these tools will depend on policies and responsibilities determined for the system.

Password Access. Access to TDOCS should be controlled through passwords and user accounts. Passwords are needed to control who does and who does not have access to the system. While user passwords could be supported by TDOCS, how they should be used is a matter of policy to be determined by the DHLWM. Alternatives range from open access to all users to restricted access to particular documents or even sets of documents.

User Privileges. One of the major reasons for having passwords is that privilege levels can be associated with them. Typically, three privilege levels are used: view, edit, and administrate. A user with view privileges only has the capability to query, search, retrieve, and view data. A user with edit privileges has the capability, in addition to view privileges, to insert, delete, and modify data. A user with administrative privileges, in addition to view and edit privileges, has the capability to define and manipulate full-text indices, database definitions, file server contents, and user accounts and privileges. Privileges are very important for maintaining confidence in the database of documents. The extent to which TDOCS implements user privileges is dependent on policies to be determined. Full-text systems, however, do not generally provide privileges for document modification but only viewing and copying.

User Accounts. User accounts are more a matter of convenience. A user account would allow an individual staff member to configure TDOCS in ways that best meet his or her needs. This might include, for example, subsets of documents, personal definitions of search concepts, private notes, or simply interface color schemes. While user accounts should be supported by TDOCS, the extent to which they could control system environments depends on the flexibility of the software selected and the needs of the staff.

3.1.5.2 Document Tracking and Configuration Control

Document tracking is important in a document management system when the documents are dynamic. While it might be important to log the entry dates for static documents, for those documents that change over time it, will be important to track versions. This becomes especially important when those documents are cited as references in reviews by DHLWM staff and used in submissions to the LSS/INFOStreams. Whether or not this is implemented in TDOCS is a matter for policy determination.

3.1.5.3 Monitoring and Reporting

In order to facilitate system administration and maintenance, TDOCS should support a variety of standard monitoring and reporting tools. These should include server and network performance monitoring, including file and database volume. These should also include descriptive reports on the document contents of the file server and database.

3.1.5.4 Backup and Recovery

Since no system, hardware or software, is impervious to natural disaster, TDOCS should provide tools to automate backup and recovery of client software and server database and files.

3.2 SYSTEM CONSTRAINTS

3.2.1 Graphical User Interface

TDOCS must be implemented using standard GUIs. These include Microsoft Windows, IBM OS/2, Sun OpenLook, and Macintosh System 7. All of the interfaces support menus, buttons, and lists for function selection that are consistent, intuitive, and easy to learn and use; multiple, scrollable windows for viewing and editing; clipboard cut and paste; and dialogues for interaction and progress feedback.

3.2.2 Acceptable Response Times

TDOCS must be implemented to provide reasonable response times for document query and search. In order to support this requirement, the full-text database should be partitioned for selective search, nonsyntactic queries and searches should be detected, search and retrieval progress feedback should be displayed, and query and search cancellation should be provided.

3.2.3 Multiple Platforms

TDOCS must be designed and implemented to support multiple platforms. A number of platforms must be considered since dual-site installation of TDOCS is required. The ACRS Pcs running Microsoft DOS/Windows and Sun Workstations running UNIX/OpenLook must be supported. At the CNWRA, at least Pcs running IBM's OS/2, Sun Workstations running UNIX/OpenLook, and MacIntoshes running System 7 must be supported.

3.2.4 Client/Server Architecture

TDOCS must be designed and implemented around the client/server architecture of the DHLWM's ACRS. Documents (text and images), indexes, and headers will likely reside in a server's file system and database. They could be accessed for distributed processing by staff at client workstations. While the aim is to minimize the impact on existing and planned systems and configurations, over time, the growing volume of documents — text and, especially, images — will drive the need for additional disk space; upgrading of the existing Sparc 10, Model 41; or possibly even the addition of a dedicated server.

3.2.5 Maximize Commercially Available, Off-the-Shelf Software

TDOCS implementation must maximize use of commercially available, off-the-shelf software (COTS) packages where prudent in design. While the specific requirements of TDOCS preclude the use of a single software package, major functional areas can be thus supported. Selection of these software packages should be based on their support for major functional areas, extensibility, adherence to industry-wide standards, provision of a suitable application program interface (API), and support for multiple platforms in a client/server architecture. Extensibility would aid in customizing the software; adherence to standards would aid in compatibility with other systems; and provision of an API would aid in the integration of packages. Any integration and extension requiring customized software should be developed using standard programming tools that likewise adhere to industry-wide standards and provide suitable APIs.

3.2.6 Impact on Existing/Planned Systems and Configurations

The impact of TDOCS on existing and planned systems and configurations, namely AUTOS and ACRS, must be minimized as much as possible. There are, however, a number of tradeoff decisions that must be made that could well require additional or enhanced hardware.

Additional server disk space may be required as the document database grows over time. As has been explained previously in Section 3.1.1, the estimated storage requirements for TDOCS should not exceed 10 gigabytes over the two year period following installation. This should have no more impact on the existing ACRS than perhaps requiring an additional external harddrive for the Sun Sparc 10. Whether or not TDOCS will incur these two additional costs depends to a large extent upon the volume of text and especially images stored. They are inevitable should policy determine a need to store images of each document in order to maintain compatibility with planned revision of NUDOCS and implementation of LSS/INFOStreams.

If display and enhancement of over-sized, high resolution, and/or 24-bit color images is desired for staff, their workstations will require enhanced video cards and monitors. In order to avoid this impact, technical staff could share an enhanced workstation located in a common area.

Access to other databases and on-line services, if not constrained, may require additional communication lines to handle data flow.

3.2.7 Expandability to Meet Evolving Needs

TDOCS must be designed and implemented to ensure expandability to meet evolving needs. Expansion is expected for access to other databases and on-line services, and eventually for coexistence with a revised NUDOCS and LSS. This sort of expansion should be supported by selecting off-the-shelf software packages that meet industry standards and, where possible, provide an API.

3.2.8 Meeting NRC Policies and Standards

TDOCS must adhere to NRC policies and standards for software development as laid out in the NRC Software Quality Assurance Program and Guidelines (NUREG/BR-0167). Thus, its implementation must follow specified life-cycle activities: requirements definition (this deliverable), design, implementation, qualification testing, installation and acceptance, operations and sustaining engineering, and retirement and archiving.

3.3 SYSTEM POLICIES

For efficient design and implementation, some significant policy matters need to be addressed. Many of these matters relate directly to maximizing user confidence.

3.3.1 Specification of "Selected" Set of Documents

In the initial phases of TDOCS document loading, emphasis should be placed on loading technical materials generated by the DHLWM, materials generated by the CNWRA, and specific materials identified or currently being used by the DHLWM staff. As loading progresses, selected sets of documents could be identified for routine loading. Policies, procedures, and responsibilities must be established so that the available loading resources are properly focused to capture the most relevant and urgently needed materials on a timely basis.

Document Types. Several types of documents are likely candidates for routine loading into TDOCS:

- Reports received by the DHLWM from the DOE
- USGS, National Laboratory, and DOE contractor reports pertaining to the Yucca Mountain Site
- Selected technical journal articles

Volume. In establishing policies pertaining to what materials should be loaded, careful consideration must be given to the volume of textual material and the number and size of the images as opposed to planned ACRS hardware expansion, the possibilities of archiving documents, and technological advances in storage media. Policy decisions will be required to properly allocate resources to support the storage and retrieval of the most important materials according to the estimates presented previously in Section 3.1.1.

3.3.2 Storage and Use of Proprietary and Copyrighted Materials.

The storage and use of copyrighted materials in TDOCS must conform to NRC internal requirements for handling proprietary materials, as well as international copyright laws regarding disclaimers and permissions, as is currently done with NUDOCS. Policies, procedures, and responsibilities must be established and enforced, especially for contractor reports and technical journal articles. TDOCS should support these policies where possible by displaying relevant notices with viewed documents.

3.3.3 Loading Procedures

Clear policies will be required to define document loading procedures and responsibilities. It is anticipated that some document loading, particularly document loading in support of immediate staff requirements, will be performed by the DHLWM secretarial staff. Routine document loading support must be arranged. A clear delineation of policies and responsibilities will greatly enhance the effectiveness of the document loading process.

On-Demand Loading. Requests for on-demand loading are expected to arise primarily from immediate staff requirements for access to specific pages of textual and/or graphic materials. In many cases, these requests are expected to be addressed by technical staff personnel who will perform their own document loading. Alternatively, the desired materials may be given to a secretary who will perform the document loading function. It is expected that the volume of materials loaded in this manner will be rather small and that selected pages rather than entire documents will be loaded. The intent of on-demand loading is to satisfy an immediate requirement by providing very rapid loading and access to the materials.

However, this approach to on-demand loading presents a problem, because the materials loaded will not represent complete documents. Therefore, partial documents will be loaded into a "temporary" area and will be replaced by the corresponding complete documents as soon as they can reasonably be loaded. The substitution of complete documents for partial documents loaded on a demand basis should be transparent to the staff. Access to documents would be accomplished through full-text or header searches, and the selected documents, whether partial or complete, would be subsequently accessed and displayed.

On-demand loading may also occur in circumstances that involve less urgency. A staff member may identify a complete document that needs to be loaded but is not included in the categories of materials selected for routine loading. Such documents will be submitted for on-demand loading by the secretarial staff. Because they are loaded as complete documents, they will be submitted for scanning, OCR, and loading in the permanent repository rather than in a temporary area if time permits.

Routine Loading. Requests for routine loading will identify groups of documents that will be obtained and loaded on a periodic or cyclical basis. For example, technical documents received by the DHLWM from DOE would be submitted for routine loading. Similarly, certain technical journals may be identified for routine loading. The routine loading activities should be designed to handle a substantial document volume. Support for routine loading of documents must be arranged. Routine loading will not address requirements for loading of partial documents or requests for immediate turnaround of high-priority materials.

Electronic Loading. Wherever possible, electronic copies of materials should be obtained and loaded to avoid unnecessary work. Such electronic full-text copies are frequently available and may be accompanied by electronic copies of the bibliographic headers. Obtaining electronic copies of documents avoids the labor and system overhead associated with document scanning and OCR activities. It also avoids the document cleanup process that can be rather resource intensive.

Duplicate Checking. Regardless of the method of loading, whether on-demand, routine, or electronic, it is particularly important to check the system prior to loading in order to determine whether the document to be loaded is already present in the system. Unless this is done, duplicates will inevitably be loaded into the system, resulting in wasted loading and storage resources. However, the impact of duplicate documents goes beyond wasted resources. When multiple copies of a document are loaded, they will all appear in response to full-text and header queries. The user will be presented with a succession of identical or highly similar entries in the selection list, and there will be no way to determine that they are actually duplicates without retrieving and examining them individually. Thus, the initial wasting of resources in the loading of duplicate documents is paralleled by the unnecessary wasting of the user's time and energy retrieving and examining the duplicate documents. Therefore, policies and procedures need to be established so that the person loading a document into the system will first perform a search to ensure that the document has not already been loaded.

Maintaining Hardcopy Files. Once a document has been loaded, it may be necessary to maintain it in a hardcopy file. As discussed above, if full-page images are not retained in the system, there will be no way to verify the authenticity of the scanned data other than by comparing it to a hardcopy document. Thus, depending on the decision about maintaining full-page images, it may be necessary to establish and maintain hardcopy files of documents to support staff reference requirements. The current policy seems to be to send all documents through DCD for storage. This is both a matter for further discussion between the DHLWM and the CNWRA and an issue that will be examined during design.

3.3.4 Retention of Images of Textual Materials

A policy decision on the retention of images of textual material will significantly impact the storage requirements for TDOCS. Initially, consideration was given to maintaining all textual information only as full text and not as images of full pages. In contrast to other planned systems, such as the LSS where all pages would be retained as images with parallel full-text as appropriate, the initial approach to TDOCS would utilize images only for those portions of the materials that could not adequately be represented by full text. Under this approach, input documents would be scanned, and textual information would be submitted for OCR conversion to full text. Images would be retained only for graphical entities, such as figures, photographs, equations, formulae, etc. This approach would optimize storage capacity because the storage requirements for full text are much less than those for images of full pages.

It has previously been estimated that, over a two year period, storing the text, images, headers, and indices of 10,000 documents would require approximately 10 gigabytes of storage space. By comparison, storing the full image of each document would increase that estimate by a ratio of 3 to 1, or an estimated 30 gigabytes of storage space for 10,000 documents in addition to the 10 gigabytes required for the full-text system. This estimate is based on the fact that, provided the images are compressed, storage would require perhaps 15 kilobytes per page. Compression would, however, slow retrieval and display time.

However, considerations associated with the verification and accuracy of TDOCS information suggest that full-page imaging may be necessary. If the full pages, including textual information, are not imaged, then there is no way to detect and adjust for scanning and document cleanup errors. For example, a scanning or document cleanup error (e.g., a mis-scanned significant digit or a manual transposition of significant digits during cleanup) could occur in a line of text that contained an important numeric parameter. A user who relied on such incorrect data in the full-text repository would have no way to verify the accuracy of the parameter other than referring to the original document or its image in TDOCS. Thus, by eliminating full-page images and only retaining images of non-textual entities, such as figures, formulae, photographs, equations, etc., the user would be deprived of an important facility for verifying and/or resolving uncertainties about the accuracy or completeness of the full-text information. Nevertheless, depending on policy decisions, staff could still have hardcopy documents available for these purposes. Moreover, many of these problems related to confidence can be avoided by obtaining electronic versions of documents where available.

The decision of whether or not to store full-page images for all scanned materials will have a major impact on both the functionality and cost of the system. Storing full-page images significantly increases system functionality and simplifies implementation of image display processes. However, it would require additional costs in document processing (scanning, OCR, and cleanup) and system administration, as well as an increase of several orders of magnitude in the storage requirements to support parallel image and full-text repositories, above and beyond the estimate presented previously in Section 3.1.1.

3.3.5 Use of Bibliographic Headers

A policy question needs to be resolved regarding the use of bibliographic headers in the TDOCS system. Many text management systems augment full-text search and retrieval capabilities with structured searches based on bibliographic headers. There are several reasons for including a bibliographic header capability in the TDOCS system.

Structured Searches. The ability to search bibliographic headers is particularly helpful when the user knows something about the documents to be retrieved. Bibliographic headers capture certain information in structured data fields. This provides a means by which users may significantly narrow a search when specific words are known to occur in one or more of these header fields. For example, a bibliographic header will normally contain the title, author, and publication date of the document along with other fields. Thus, a user may search bibliographic headers to retrieve all documents written by a particular person, on a particular subject, and/or during a specified period of time.

Concise Definition of Document Subject. Some bibliographic headers include fields, such as subject descriptors, keywords, etc., that provide a concise method of searching for specific document content. When properly prepared and controlled by a thesaurus, these bibliographic header fields can be very effective, because the subject descriptors are predictable. The user does not have to guess about significant search terms because the desired document may be retrieved using clearly defined and consistent subject descriptors, even if those descriptors do not appear in the text of the document. How important extended headers are for TDOCS is a matter for design since the use of full-text indexing more or less obviates the need for any more than standard bibliographic citations.

Determination of Relevancy. Bibliographic headers provide a convenient method for a user to view pertinent information about a number of selected documents and determine which ones are likely to be relevant. When a search is performed, the resulting selection list may be extensive. Scanning the bibliographic headers of selected documents can help a user to quickly focus on the desired documents.

Description of Non-Textual and Non-Imagable Materials. Bibliographic headers provide an effective means of describing materials that cannot be identified through full-text searches. For example, a geologic map would be captured as an image, but the OCR process would not capture any meaningful text. Therefore, the important information about maps must be captured in bibliographic headers to permit users to search for and retrieve these materials. Similarly, non-scannable materials, such as data on magnetic tapes, must be described by bibliographic headers.

There are several approaches that may be taken to prepare and maintain bibliographic headers. These are preparation of headers by professional staff, use of catalogers to prepare headers, and preparation of headers by clerical staff. These approaches differ both in the quality of the bibliographic headers produced and in the level of effort required to support and maintain them.

Preparation of Headers by Professional Staff. When documents address highly technical subject matter, it may be appropriate to use professional staff to prepare the bibliographic header. For example, the author of a document is intimately familiar with its content and is in the best position to describe that document in the bibliographic header. Similarly, the technical staff member submitting a document for scanning should be able to describe the content of that document accurately in a bibliographic header. However, technical staff members are not usually motivated to perform the data preparation and data entry tasks required to create a bibliographic header, and considerations of cost and loss of professional staff productivity all but rule out this approach. Thus, the technical staff, who are best equipped to describe materials in bibliographic headers, are not usually involved in this process.

Use of Catalogers to Prepare Headers. An alternative approach is to use professional catalogers to prepare bibliographic headers. This approach usually produces high-quality and consistent bibliographic headers, but the cost of the catalogers can be quite high when considered against the benefit derived from the headers. Thus, professional catalogers are usually employed only for very large document management systems that use very extensive and complex bibliographic headers.

Preparation of Headers by Clerical Staff. A more cost effective approach involves using clerical staff to prepare bibliographic headers based on information that is readily available in the first few pages of the document. This approach requires that the bibliographic header be structured to include only those fields, such as title, abstract, author, date, publisher, journal name, volume, issue, report number, etc., that are easily identified and entered by clerical personnel. The disadvantage of this approach is that the content of the bibliographic header will be more limited and subject descriptors and keywords will not be available to the user.

3.3.6 Confidence and Quality Assurance

Policies and procedures must be established that encourage user confidence in the system and that support any quality assurance (QA) requirements. From a QA perspective, a record is not a record until it has been validated. However, conversion of validated records, either by machine or manual methods, effectively invalidates them. Thus, depending on the level of QA considerations that are deemed appropriate to the TDOCS system, it may be necessary to perform some type of validation following the scanning, OCR, and cleanup processes.

Scanning, OCR, and Cleanup. Document scanning and OCR are automated processes that are normally quite accurate and reliable. Poor quality input documents, however, can have a very adverse effect on the quality of the scanning and OCR processes and can induce errors that will require document cleanup. Document cleanup is normally a manual process requiring an operator to view the image and/or the original document and make appropriate corrections. In some cases, errors will be detected automatically. For example, when characters are sufficiently distorted or blurred that the OCR process cannot recognize the character unambiguously, an error condition will be signaled. However, some errors will not be detected automatically. For example, a malformed "G" could easily be interpreted as a "C." In such cases, the error would not be detected by the OCR process, but it might be detected by some other method, such as spell checking. However, even such validation techniques might not be sufficient in certain cases. For example, if a "G" in the word "GRAY" were misread as a "C," the resulting word, "CRAY," could pass some spell checkers.

Headers. In some cases, headers will be downloaded from other systems, such as NUDOCS, and converted to the format used in TDOCS. In such cases, the header content could be assumed to be accurate. Most headers, however, will be entered manually. This introduces the possibility of many types of errors in the header fields, including misspellings, transpositions, omissions, etc. At a minimum, headers should be visually reviewed and validated. Some limited automated validity tests may also be performed during header entry to ensure that technical constraints (e.g., length and format of date fields, numeric information not allowed in author name field, etc.) are observed.

Correction of Errors. The system needs to provide a mechanism for correcting errors when they are detected by users. If a user finds a significant error in a document, then the error should be corrected to prevent other users from having to deal with it later. This is a relatively simple process in the case of textual data, but it is much more complicated to correct an error in an image. In general, defective images would be re-scanned and replaced. A more likely type of error associated with images would be when an image file was misidentified or linked to an inappropriate page of text. This type of error would be detected by a user when an inappropriate image is displayed from an icon or hyperlink in the full-text display. Corrections of such errors can be extremely difficult and should probably be addressed by re-scanning the document and submitting it again for OCR and loading.

3.3.7 System Administration, Maintenance, and Training

Administration and maintenance will require a set of policies concerning procedures and responsibilities. Policies must be set up to control user access and privileges. It must be determined who may access the system and what privileges each user has when using it. It has been suggested that three privilege levels be instituted — view, edit, and administrate levels.

Administration. Administrator responsibilities must be clearly specified. The administrator will likely maintain TDOCS, assign passwords and privileges, and possibly assist in the use of scanners, OCR and cleanup software. Responsibilities related to TDOCS must be determined with consideration to other responsibilities and workload that the administrator may have for other systems.

Maintenance. The same policy for overall ACRS hardware/software maintenance will be used.

Training. The same policy for overall ACRS training responsibilities will be used.

3.3.8 DHLWM/CNWRA Synchronization

Since installation of TDOCS at DHLWM and CNWRA sites is a requirement, policies, procedures, and responsibilities must be determined for synchronization. Alternatives include the following:

- Single repository maintained at one site, one site accessing via a LAN, the other via communication lines
- Both sites load, index, and header data into respective repositories; data shared by tape or optical disk
- All loading, indexing, and headers done at one site; other site updated via tape or optical disk

The first alternative is an attractive ideal but may not be realistic given the current configuration of the NRC/CNWRA communications systems; it remains as a future upgrade. The second alternative presents problems coordinating, especially, on-demand loading simply because both sites may request the same document. The third alternative avoids duplication but introduces delay. Maintenance of a central file of headers for on-line duplicate check would usefully augment any of these approaches. Determination of which approach is to be taken is a matter of policy.

3.4 SUMMARY OF REQUIREMENTS

By way of summary, Figure 3-1 diagrams the major functional requirements of TDOCS. Whereas previous sections of this requirements definition have taken an analytical approach, the figure provides a synthesis of system functionality. While many details, such as system administration and maintenance, constraints and policies have been omitted, it does depict in a succinct way the intended sources and flow of technical documents through the system. While supported by some software tools, system administration and maintenance involve policies, procedures, and responsibilities the DHLWM needs to enforce to ensure immediate and continued confidence in and usefulness of TDOCS. Table 3-1 also summarizes TDOCS requirements. It considers functions, constraints, and policies in terms of support by commercial software, required customization, required policy decisions, and dependency on schedule or the facilities of other systems.

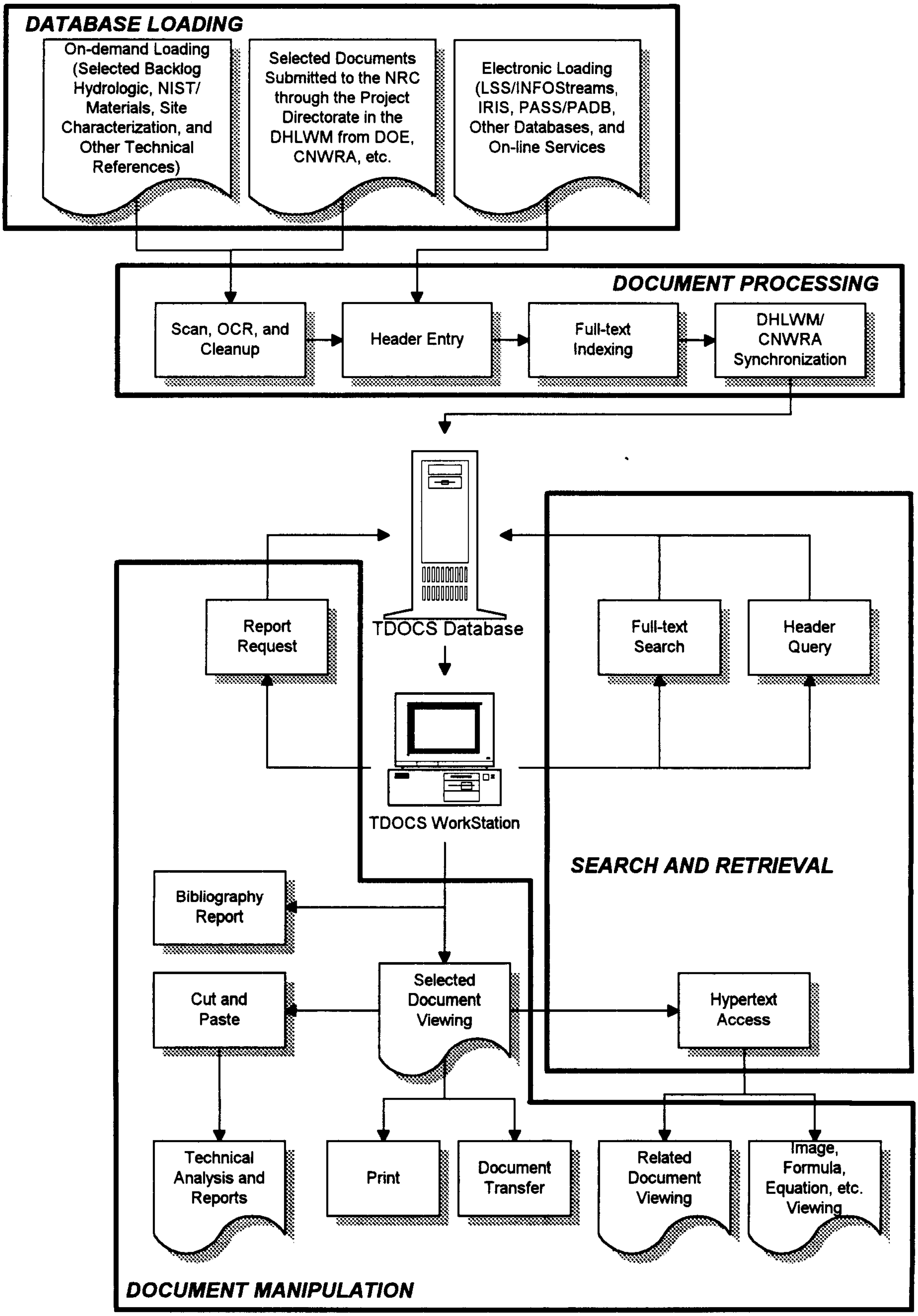

DATABASE LOADING
On-demand Loading (Selected Backlog Hydrologic, NIST/ Materials, Site Characterization, and Other Technical References)
Selected Documents Submitted to the NRC through the Project Directorate in the DHLWM from DOE, CNWRA, etc.
Electronic Loading (LSS/INFOStreams, IRIS, PASS/PADB, Other Databases, and On-line Services
DOCUMENT PROCESSING
Scan, OCR, and Cleanup
Header Entry
Full-text Indexing
DHLWM/ CNWRA Synchronization
TDOCS Database
Report Request
Full-text Search
Header Query
TDOCS WorkStation
SEARCH AND RETRIEVAL
Bibliography Report
Cut and Paste
Selected Document Viewing
Hypertext Access
Technical Analysis and Reports
Print
Document Transfer
Related Document Viewing
Image, Formula, Equation, etc. Viewing
DOCUMENT MANIPULATION

Table 3-1. Summary of TDOCS Requirements.

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
DL	DATABASE LOADING				
DL-1	Routine loading	Scanning and OCR supported by COTS.	Custom code required to load text and images into repositories.	Definition of responsibilities. Selection of document groups for routine loading.	Depends on availability and quality of source documents. Electronic loading depends on formats and interfaces required by other systems
DL-2	On-demand loading	Scanning and OCR supported by COTS.	Custom code required to load data into temporary areas and then replace the temporary entries when complete documents are loaded.	Definition of responsibilities. Procedures for loading and replacement of partial documents.	Depends on availability and quality of source documents.
DL-3	Electronic loading from tapes, disks, etc,; other databases, such as NUDOCS, PASS/PADB,and IRIS; and on-line bibliographic services	File transfer may be supported in part by COTS. Communications facilities should be supported by COTS.	Extensive custom code may be required to reformat and load documents received on electronic media and to interface with other systems, reformat, and load documents.	Definition of systems and document sources to be used. Definition of systems to be interfaced with TDOCS. Definition of download procedures and institutional agreements.	Incompatibilities between formats and may affect the ability to achieve document loading from magnetic media. Incompatibilities between systems, formats, and protocols may affect ability to download. Availability of data and system interfaces. Implementation schedules for different systems.
DP	DOCUMENT PROCESSING				
DP-1	Scanning, OCR, and Cleanup	Scanning and OCR supported by COTS.	Custom code required to support cleanup and loading of documents.	Definition of approach to cleanup of scanning and OCR errors.	N/A

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
DP-2	Bibliographic header entry	Little support provided by COTS.	Custom software required to support entry of headers and formatting of headers to support full-text search.	Decision to include headers in TDOCS. Decision on level of compatibility required between headers for TDOCS, NUDOCS, LSS, and other systems.	Compatibility with other systems may affect requirement for catalogers and complexity of entry/maintenance software.
DP-3	Full-text indexing	Supported by COTS.	Custom software required for pausing.	Decision to support header searches through full-text facilities.	N/A
SR	SEARCH AND RETRIEVAL				
SR-1	Document access: full-text search	Fully supported by COTS.	N/A	Decision to use slip sheets to represent non-text-searchable materials.	N/A
SR-2	Document access: structured header queries	Supported by COTS if implemented through full-text search facilities.	Requires custom software to format headers so that they can be searched by the full-text system.	Decision to support header searches through full-text search facilities. Decision to use slip sheets to represent non-text-searchable materials.	N/A
SR-3	Document access: hyperlinks	Supported by COTS.	May require custom code to support automatic generation of hyperlinks for images.	Decision on private/public hyperlinks.	N/A
SR-4	Search confidence: wildcards & Boolean operators	Supported by COTS.	N/A	N/A	N/A
SR-5	Search confidence: phrase search	Supported by COTS.	N/A	N/A	N/A
SR-6	Search confidence: near-spell search	Supported by COTS.	N/A	N/A	N/A
SR-7	Search confidence: fuzzy search	Supported by COTS.	N/A	N/A	N/A

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
SR-8	Search confidence: proximity search	Supported by COTS.	N/A	N/A	N/A
SR-9	Search confidence: search result ranking	Supported by COTS.	N/A	Procedures for relevancy weighting of search results.	N/A
SR-10	Search confidence: cross-partition search	Supported by COTS.	N/A	Policy on partitioning and organization of documents.	N/A
SR-11	Concept search and building	Supported by some COTS.	N/A	Decision to place emphasis on concept-based searches. Decision on private/public concept-based searches. System administration policies on creation and maintenance of concept definitions.	N/A
SR-12	Query save, recall, and edit	Supported by COTS	N/A	N/A	N/A
SR-13	Search result browsing	Supported by COTS.	N/A	N/A	N/A
SR-14	Concurrent, multiple document viewing and scrolling	Supported by COTS.	N/A	N/A	N/A
SR-15	In-document hit highlighting and browsing	Supported by COTS.	N/A	N/A	N/A
SR-16	Hyperlink creation	Supported by COTS.	N/A	System administration policies on who can create hyperlinks between documents and how they should be managed.	N/A

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
DM	DOCUMENT MANIPULATION				
DM-1	Cut, copy, and paste	Supported by COTS.	N/A	N/A	N/A
DM-2	Document printing	Supported by COTS. Printing of images may have restrictions.	Custom software and/or integration may be required for images and special formats.	N/A	N/A
DM-3	Report generation (NIST/Materials database)	May be supported by database COTS.	Requires custom software and/or report specifications through COTS report generation languages.	Decision on extent and nature of reporting functions	N/A
DM-4	Document transfer (e-mail)	Transfer supported by COTS.	Custom software may be required to extract documents from the repository and interface to e-mail.	Decision on file transfer procedures versus access.	Formats between TDOCS and file recipient's environments may be incompatible.
DM-5	Public and private notes	Supported by COTS.	N/A	Decision on public/private notes.	N/A
DM-6	Image manipulation and enhancement	Supported by COTS.	N/A	Decision on need for image manipulation facilities. Assessment of hardware/software cost and impact.	N/A
AM	ADMINISTRATION AND MAINTENANCE				
AM-1	Administration and maintenance: password access	Supported by COTS.	N/A	Decision on level of security to be implemented through passwords.	N/A

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
AM-2	Administration and maintenance: user privileges	Supported by COTS.	Some custom software may be required to implement restrictions on view, update, administrate privileges.	Decision on classes of users and their respective privileges. Definition of loading privileges and procedures. Definition of system administration responsibilities.	N/A
AM-3	Administration and maintenance: user accounts	Supported by COTS.	N/A	Decision on whether and/or how to segment information for user access.	N/A
AM-4	Document tracking and configuration control	N/A	Custom software required.	Decision on requirement for and level of configuration control. Decision on retention of superseded documents.	N/A
AM-5	Monitoring and reporting	Limited COTS support.	May require custom software.	System administration policies and responsibilities.	N/A
AM-6	Backup and recovery	Supported by COTS for relational repositories, by operating system for documents.	Custom software required to automate.	System administration policies.	N/A
SC	SYSTEM CONSTRAINTS				
SC-1	Graphical user interface	Supported by COTS.	Must be supported in custom software.	N/A	N/A
SC-2	Acceptable response times	N/A	Custom software must support response feedback.	Definition of acceptable response time.	N/A
SC-3	Multiple platforms	Supported by COTS.	Must be supported by custom software.	Definition of platforms to be supported.	N/A
SC-4	Client/server architecture	Supported by COTS.	Must be supported by custom software.	N/A	N/A

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
SC-5	COTS	Available for scanning, OCR, full-text, relational database, communications, and E-mail functions.	Custom software will be needed to interface COTS.	N/A	N/A
SC-6	Impact on existing/planned systems and configurations	N/A	N/A	Clear definition of planned hardware/software and application environments.	Required system interfaces may be difficult to implement. Schedule for implementation of systems. Hardware limitations.
SC-7	Expandability to meet evolving needs	Supported by COTS.	Requires careful design and implementation of custom software.	Definition of future requirements. Possible scoping of repository sizes and document loading volumes.	Requirements for system interfaces with NUDOCS, and when available, the NUDOCS revised system, IRIS, and LSS/INFOStreams.
SC-8	Meeting NRC policies and standards	N/A	N/A	Determination of policy.	N/A
PC	SYSTEM POLICIES				
PC-1	Specification of "selected" set of documents	N/A	N/A	Selection of document set.	Compatibility and interfaces with other systems.
PC-2	Storage and use of proprietary and copyrighted materials	N/A	N/A	Recognition of rules and laws.	N/A
PC-3	Loading procedures	See Database loading.	See Database loading.	Determination of procedures.	N/A
PC-4	Retention of formatting and images	Basic scanning facilities supported by COTS.	Custom code required to select images and associate them with text.	Definition of approach to full-page imaging.	Incompatibility of image formats with other systems may affect ability to retain formatting and images.

ID	REQUIREMENTS	SUPPORTED BY COTS	REQUIRES CUSTOM CODE	REQUIRES POLICY DECISION	DEPENDENT ON SCHEDULE OR FACILITIES OF OTHER SYSTEMS
PC-5	Use of bibliographic headers	See Bibliographic header entry.	See Bibliographic header entry.	Determination of policy and procedures.	N/A
PC-6	Confidence and quality assurance	See Search and retrieval.	See Search and retrieval.	Determination of policy and procedures.	N/A
PC-7	System administration, training, and maintenance	Limited COTS support.	May require custom software.	System administration policies and responsibilities.	N/A
PC-8	DHLWM/CNWRA synchronization	N/A	Requires custom software.	Policy on synchronization timing and procedures.	Interfaces between TDOCS, PASS/PADB, and other systems.

4 CONCLUSIONS AND DIRECTIONS

The purpose of this report is to identify overall requirements for TDOCS and facilitate the analysis and decision making necessary to initiate its design. The system must meet the specific and real needs of the DHLWM staff. In order to implement a system that is immediately useful, system requirements must be driven by the notion of confidence. Confidence, defined as recall (finding all relevant documents) and precision (finding only relevant documents), is a constraint on all of the highly interrelated functional aspects of TDOCS listed above. The notion of confidence in document management systems like TDOCS must be understood and appreciated in order to make appropriate decisions and commitments regarding impact, cost, and tradeoffs in the system's usefulness. These decisions and commitments must then be reinforced with policies, procedures, and responsibilities clearly defined by the DHLWM for using TDOCS, so that it will not only provide immediate benefit but evolve with the needs of its users.

Section 2 described the technical document reference database needs of a number of related systems, such as, the LSS, DOE's INFOStreams and IRIS, NRC's NUDOCS and Office databases, the ACRS, and the CNWRA's PASS/PADB. It was found that, for TDOCS, the most unique requirements and pressing policy and compatibility issues driven by these systems are support for:

- Limited capabilities for the DHLWM with an interface in the future to the LSS for official HLW documents
- Header and full-text search
- Images of non-textual material, such as formulae, equations, tables, graphs, charts, pictures, and photographs
- On-demand scanning to meet immediate needs
- Incorporation of materials and references in analyses
- Connectivity with other systems and databases

The overall set of specific requirements for TDOCS were analyzed in Section 3 in terms of applicable functions, constraints, and policies. They include the following:

- Database loading: getting documents into the database
- Document processing: cleaning up, entering headers, and indexing full text
- Search and retrieval: accessing the documents
- Document manipulation: facilitating document use
- Administration and maintenance: ensuring system functionality

This section concludes the report by summarizing proposed and implied requirements that were confirmed for TDOCS by this analysis. It then provides directions to be taken for resolving open issues for system design and policy determination.

4.1 REQUIREMENTS CONFIRMED

The following requirements were confirmed as TDOCS functions and constraints. These address the complete set of requirements initially provided by the DHLWM or implied in that request. While all of the following requirements were confirmed, many involve open-ended issues presented subsequently as either technical issues that can only be resolved during system design or policy matters to be determined in discussions between the DHLWM and the CNWRA at the appropriate time.

4.1.1 Proposed Functions

- On-demand text and image scanning, OCR, indexing, storage, full-text search, and retrieval of selected documents needs to be supported. Quality of scanning, OCR, and preservation of formatting and images remain as technical issues for system design. Loading documents, capturing images, assuring quality of cleanup, and entering bibliographic headers are matters for policy determination.

- Routine scanning and loading of a selected set of technical documents submitted to the NRC relevant to HLW program for full-text search and retrieval needs to be supported. Quality of scanning, OCR, and preservation of formatting and images and support for access confidence remain as issues for system design are issues for system design. Specification of a "selected" set of documents, as well as procedures and responsibilities for loading documents, capturing images, assuring quality of cleanup, and entering bibliographic headers remain as matters for policy determination.

4.1.2 Proposed Considerations

- Consolidation of three in-house technical databases needs to be supported.

- Incorporation of text and images, including color, in technical analyses needs to be supported.

- Access to documents in other databases, such as NUDOCS and PASS/PADB, needs to be supported. How these databases might be accessed is an issue for system design. Access to other databases is a matter for policy determination.

- Dial in to on-line library bibliographic services needs to be supported. How these services might be accessed is an issue for system design. Which services might be supported and the purpose for dial-in are matters for policy determination.

- Transfer of documents and, possibly, technical data packages from DOE's INFOStreams needs to be supported with emphasis primarily on individual documents and only secondarily on packages. How these documents and packages might be transferred and stored is an issue for system design. The extent to which this might be done is a matter for policy determination.

4.1.3 Proposed Constraints

- GUIs need to be supported.

- Multiple platforms need to be supported.

- Client/server architecture needs to be supported.

- COTS packages that adhere to standards and, where possible, provide an API, need to be used. The need for customizing such software and for developing custom code is an issue for system design.

- Existing/planned ACRS needs to be supported. Possible expansion of ACRS to meet the computing and storage needs of TDOCS is an issue for system design

- Expandability to meet evolving needs should be supported.

4.1.4 Implied Constraints

- Compatibility of design with CNWRA systems and synchronization of TDOCS at both sites need to be supported. How this might be accomplished is an issue for system design. Procedures and responsibilities for synchronization is a matter for policy determination.

- The loading of electronic documents needs to be supported. The sources and volume of such documents has been estimated but may be a matter for policy determination.

- Development of policies for system administration and maintenance need to be supported with COTS customization. Specification of these policies is a matter for policy determination.

- Both structured header queries and full-text search need to be supported.

- Confidence in document database scope and access needs to be supported. The degree to which confidence is ensured through selection of document, accurate cleanup, and complete header entries are matters for policy determination.

- Viewing, printing, and transferring documents; cutting and pasting; storing references, embedding notes; and creating hyperlinks need to be supported. The degree to which the system facilitates these utilities is an issue for system design.

- NRC policies and standards need to be adhered to.

4.2 ISSUES FOR SYSTEM DESIGN

The following issues represent technical tradeoff decisions that can only be effectively decided on during the design of TDOCS.

- Quality of scanning, OCR, and preservation of formatting and images, that is, the degree to which foreseeable problems can be overcome, need to be determined during design and will involve possible tradeoff decisions in the selection of software packages.

- The degree to which query and search mechanisms are supported need to be determined during design and will involve possible tradeoff decisions in the selection of software packages.

- The degree to which the system facilitates user-created hyperlinks, public/privates notes, and search concepts need to be determined during design and will involve possible tradeoff decisions in the selection of software packages.

- The need for customizing off-the-shelf software and for developing custom code, particularly for integration, needs to be determined during design and will involve possible tradeoff decisions in the selection of software packages.

- Expansion of ACRS, particularly to handle data volume and imaging requirements, needs to be determined during design but is closely tied to matters for policy determination.

4.3 MATTERS FOR POLICY DETERMINATION

The following matters remain to be determined as policies, procedures, and responsibilities in discussions between the DHLWM and the CNWRA, and in some cases may require participation of IRM. For effective design of TDOCS to take place these matters must be resolved as soon as possible. They include:

- Routine loading: Definition of responsibilities and specification of selected document groups remain for policy determination between the DHLWM and the CNWRA.

- On-demand loading: Definition of responsibilities and specification of selected document groups remain for policy determination between the DHLWM and the CNWRA.

- Electronic loading: Definition of document sources to be used, systems to be interfaced with, and download procedures and institutional agreements remain for policy determination among the NRC's IRM, the DHLWM, and the CNWRA.

- Scanning, OCR, and cleanup: Definition of approach to cleanup of scanning and OCR errors remains for policy determination between the DHLWM and the CNWRA.

- Bibliographic header entry: Decisions to include headers and the level of compatibility with other systems remain for policy determination between the DHLWM and the CNWRA.

- Full-text indexing: Decision to support header searches remains for policy determination between the DHLWM and the CNWRA.

- Document access: Decisions to use slip sheets for non-text-searchable materials, to support header searches, and to provide private/public hyperlinks remain for policy determination between the DHLWM and the CNWRA.

- Search confidence: Decisions on partitioning and organizing documents, placing emphasis on concept-based searches, constructing public/private concept definitions, and weighting the relevancy of search results remain for policy determination between the DHLWM and the CNWRA.

- Hyperlink creation: Decisions on the administration and management of hyperlink creation remain for policy determination between the DHLWM and the CNWRA.

- Report generation: Decisions on the extent and nature of report functions remain for policy determination between the DHLWM and the CNWRA.

- Document transfer: Decision on file transfer procedures versus access remains for policy determination between the DHLWM and the CNWRA.

- Public and private notes: Decision on public and private notes remains for policy determination between the DHLWM and the CNWRA.

- Image manipulation and enhancement: Decision on the need for image manipulation facilities and assessment of hardware/software cost and impact remain for policy determination between the DHLWM and the CNWRA.

- Administration and maintenance: Decisions concerning levels of password security, user accounts and privileges, and administrative procedures and responsibilities remain for policy determination between the NRC's IRM, the DHLWM and the CNWRA.

- Document tracking and configuration control: Decisions on requirements for configuration control and retention of superseded documents remain for policy determination between the DHLWM and the CNWRA.

- Monitoring and reporting: Decisions on support for administration policies and responsibilities remain for policy determination between the NRC's IRM, the DHLWM, and the CNWRA.

- Backup and recovery: Decisions on support for backup and recovery remain for policy determination between the NRC's IRM, the DHLWM, and the CNWRA.

- System constraints: Definition of acceptable response times, platforms to be supported, hardware/software platforms, future requirements, and NRC software requirements remain for policy determination between the NRC's IRM, the DHLWM, and the CNWRA.

5 REFERENCES

Advanced Management Systems, Inc. 1991. *Requirements Analysis of Text Management Needs for the Nuclear Regulatory Commission.* Rockville, MA: Advanced Management Systems, Inc.

Advanced Management Systems, Inc. 1993. *ACRS/ACNW Text Management and Imaging System: Scope and Requirements.* Rockville, MA: Advanced Management Systems, Inc.

Advanced Management Systems, Inc. and Pinkerton Computer Consultants, Inc. 1992. *Draft Final Report of the Text Management Prototype Project.* Rockville, MA: Advanced Management Systems, Inc.

Chery, D.L. 1990. DHLWM computer hardware and software functional needs and some proposed specific needs. *Report of the Task Group for Evaluation of the Division of High-Level Waste Management Computer Hardware and Software Needs.* Washington, DC: U.S. Nuclear Regulatory Commission.

Chery, D.L. 1992. Task 6: DHLWM Advanced Computer Review System. Enclosure: "Text Management Prototype Project Overview." Note to Rawley Johnson, Center for Nuclear Waste Regulatory Analyses. Washington, DC: U.S. Nuclear Regulatory Commission.

Cerny, B.A., 1992. Information Management for the Department of Energy Office of Civilian Radioactive Waste Management. *Proceedings of the Third Annual International Conference on High Level Radioactive Waste Management, April 12-16.* La Grange, IL: American Nuclear Society, Inc. and New York, NY: American Society of Civil Engineers.

Chilk, S.J. 1993. SECY-93-107 — Licensing Support System Program and Budget Responsibilities. Memorandum for James M. Taylor, Executive Director of Operations. Washington, D.C.: U.S. Nuclear Regulatory Commission.

Center for Nuclear Waste Regulatory Analyses. 1993. *(DHLWM) Advanced Computer Review System for Technical License Review — Support Tasks.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses (CNWRA).

Center for Nuclear Waste Regulatory Analyses. 1993. *CNWRA FY94-95 Operations Plan for the Division of High-Level Waste Management.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

DeWispelare, A.R., J.H. Cooper, R. D. Johnson and R. L. Marshall. 1992. *Review and Analysis of the PASS/PADB System for Systematic Regulatory Analysis.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

DeWispelare, A.R., R.D. Johnson, R. L. Marshall, and J. H. Cooper. 1993. *Development Plan for PASS/PADB System Design Version 3.0.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Gianios, C. 1991. *AUTOS — Detailed Design Plan and Inventory of DHLWM Computers.* Washington, DC: U.S. Nuclear Regulatory Commission.

Harloe, E. 1993. *External Database Access Options Report.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Hoyle, J.C., Chairman, LSS Advisory Review Panel, USNRC. 1993. Correspondence to H. W. Swainston, Deputy Attorney General, State of Nevada.

Johnson, R.D., S.R. Young, C. L. Acree, Jr., and J. H. Cooper. 1991. *Alternative Ways of Making Packaged Documentary Materials Accessible within the Licensing Support System.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Johnson, R.D., R.L. Marshall, and S. W. Dellenback. 1992a. *Functional Needs Update and Status Report on the DHLWM Advanced Computer Review System.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Johnson, R.D., R.L. Marshall, S. W. Dellenback, and R. H. Martin. 1992b. *Design and Implementation Plan for the DHLWM Advanced Computer Review System.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Johnson, R.D., and R.L. Marshall. 1992. *Proposed Analysis and Design Tasks for the DHLWM Advanced Computer Review System.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Johnson, R.D., and C. Moehle. 1993. *Meeting Report on Defining TDOCS Requirements.* San Antonio, TX: Center for Nuclear Waste Regulatory Analyses.

Meehan, B. 1993. Addition of Two New Subtasks Under Center Operations Element, Task 6 of the HLW FY93/94 Operations Plan Under Contract No. NRC-02-88-005. Correspondence to W. C. Patrick, President, CNWRA. Washington, DC: U.S. Nuclear Regulatory Commission.

NRC Executive Director (EDO). 1991. *Staff Expertise and Capabilities to Utilize Analytical Codes.* SECY-91-247. Washington, DC: U.S. Nuclear Regulatory Commission.

Sheats, D.G. 1992. Records management in support of the licensing process for the high-level radioactive waste facility. *Proceedings of the Third Annual International Conference on High Level Radioactive Waste Management, April 12-16.* La Grange, IL: American Nuclear Society, Inc. and New York, NY: American Society of Civil Engineers.

Youngblood, B.J. 1992. Implementation of Document Control Procedure for Division of High-Level Waste Management. Memorandum to Division of High-Level Waste Management Staff. Washington, DC: U.S. Nuclear Regulatory Commission.

Youngblood, B.J.. 1993. Overall Review Strategy for the Nuclear Regulatory Commission's High-Level Waste Repository Program. Note to Division of High-Level Waste Management Staff. Washington, DC: U.S. Nuclear Regulatory Commission.

APPENDIX

Table A-1. AMS prototype study requirements (text quoted from AMS, 1991)

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R1	General Requirements	The system shall provide capabilities that support a set of general requirements.		
R101	Ability to accept full-text documents	The system must have the ability to accept full-text documents in electronic form (either flat ASCII file formate or as a structured record).	F	F
R102	Remote access to stored data	The system shall provide full-text and indexes search capabilities through remote dial-up access.	F	F
R103	Electronic request for paper copies	Users of the system shall be able to file an electronic request for paper copies of documents from their work stations.	N	F
R104	Search access through both indexes and full text	The system shall provide access to documents through both structured index searching (on header fields) and through full-text searching.	F	F
R105	Access control through passwords	User names and passwords shall be used to enforce access restrictions. Access to the system shall be controlled and restricted to authorized users.	F	F
R106	Updating limited to authorized users	The system shall include features to prevent unauthorized access and willful or accidental damage to the data base contents by misuse, and to limit updating to only authorized users.	F	F
R107	Direct search output to screen or printer	The system shall be capable of delivering output from searches on video display terminals and on printers. The requesting of a document to be printed and delivered, can be invoked any time during the viewing of documents.	N	F
R108	E-Mail capability	E-mail systems shall be provided with password security access and identification which will provide privacy on the contents of a mailbox as well as authentication of the sender's identity. The System Administrator shall be able to broadcast electronic messages to every authorized user of the system, and shall be able to receive electronic messages from every authorized user of the system. These capabilities may be provided by the system or by an interface to an agency-wide E-Mail system.	N	

F=full support, P=partial support, N=no support, blank=unknown

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R109	Backup and recovery facilities	They system shall support the data base administrator by providing the capability to perform routine backups of the data, execute recovery procedures when required, facilitate storage and maintenance of backups (both on-site and off-site), and execute various check routines to ensure the physical integrity of the data base.	F	F
R2	Query/Search	The system shall provide capabilities that support user queries and searching of the system data base.		
R201	Search on field(s), text, or any combination	A query shall be applicable to a single field, multiple fields, all fields, text, text and any combination of header fields, as specified.	F	F
R202	Wild card capability	A wild card capability (e.g., "document*" for "document," "documents," or "documentation") shall be available in all query construction modes.	F	F
R203	Support for Boolean operators	The use of Boolean operators (e.g., AND, OR, etc.) shall be supported.	F	F
R204	Proximity searches must be supported	Proximity (such as, within N words, within N sentences, within N paragraphs) shall be available for both header fields and full text.	N	F
R205	Phrase searches must be supported	Phrase searching shall be available for both header fields and full text.	F	F
R206	Easy-to-use cross partition (data set) searching	An easy-to-use way to apply a query to a set of data partitions shall be provided (if applicable).	N	F
R207	Feedback on search progress or expected time	Indication as to what the relative response time will be for a given query, as well as to the progress of the search, shall be provided.	P	F
R208	Detection of "unreasonable" searches	The system shall be able to detect and prevent "unreasonable" queries which would overburden the system.	N	F
R209	Off-peak scheduling of demanding queries	The system shall detect and schedule particularly demanding queries for off-peak hours and inform the user of the action taken.	N	F

F=full support, P=partial support, N=no support, blank=unknown

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R210	On-screen selection of pointer for direct access	The system shall provide on-screen selection of a pointer to directly retrieve and display the associated header or document.	N	F
R211	Save user queries for future use	The system shall provide the capability to save individual user queries for use in the same or future sessions.	F	F
R212	User ability to cancel queries in progress	A user shall be able to cancel a query both during construction and while waiting for the response.	F	F
R213	On-line display of thesaurus	There shall be an on-line display of a thesaurus that will show broader or narrower terms and related terms to assist with the selection of descriptor search terms.	F	F
R214	Capability of editing prior queries	The system will provide the capability of making simple modifications (editing) of prior queries without having to retype the entire query.	F	F
R215	Concatenation of queries with back-out feature	The system shall provide the capability for concatenation of queries, with the ability to step back if a newly concatenated query results in too small or too large a result set.	P	F
R216	Ability to examine index portions	There shall be an ability for a user to examine a segment of the index.	N	F
R217	Full text stored, indexed, and searched on-line	The full text of selected documents shall be stored, indexed, and made available for on-line search and retrieval.	N	F
R218	Full text searches of document and header	Full text search capability on both document text and headers shall be available.	F	F
R219	Keywords assigned by a controlled vocabulary	The system shall provide the capability to assign subject terms and keywords either manually or with the aid of a controlled vocabulary.	F	F
R220	Structured index searching capability	Structured index searching via detailed and extensive headers shall be available, involving subject terms and keywords.	N	F

F=full support, P=partial support, N=no support, blank=unknown

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R221	Near spell and synonym search capability	Both near spell (such as suffixes and plural endings) and synonym search capability shall be provided.	F	F
R3	Retrieval	The system shall provide capabilities that support the retrieval and display of data base records.		
R301	References to non-document data are needed	The system shall provide the capability to reference non-document data (exhibits, samples, etc.).	N	F
R302	Bookmark capability for later viewing	The capability to mark a document from various header displays, for subsequent viewing of ASCII text, shall be provided.	N	F
R303	Ability to rank hits and sort relevant headers	The system shall provide the option to rank hits and display headers in relevant order.	N	F
R304	Header displays sortable by any set of fields	Header displays shall be sortable by any set of header fields. The articles (The, A, An) should be removed for sorting titles and organizations.	N	F
R305	Highlighted query terms during display	Only the terms used in the query should be highlighted as hits in both header and ASCII text displays.	N	F
R306	Capture or download ASCII text and headers	The system should be able to support transmission of ASCII data including both text content and headers for downloading.	F	F
R307	User-selected header field display	They system shall provide the user the ability to select a subset of the header fields to be displayed during retrievals.	N	F
R308	Brief header field display mode	The system shall provide a brief display mode that displays a pre-established subset of header fields.	F	F
R309	Full header field display	The system shall provide a display mode that shows all header fields for each record retrieved.	F	F
R310	Full page ASCII text display	The system shall provide the capability to display a full page of a retrieved document on work stations equipped for that type of display.	N	F

F=full support, P=partial support, N=no support, blank=unknown

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R311	Full screen ASCII text display	The system shall provide a display mode of a full work station screen of ASCII text at a time.	F	F
R312	Key Word In Context (KWIC) display	The system shall provide a Key Word in Context (KWIC) display mode.	N	F
R313	Counts of retrieved items shall be provided	Summary descriptions of retrieval information and counts of retrieved items shall be provided.	F	F
R314	Support for linkages to image/microfilm	The system shall support linkages between retrieved text and corresponding page images or microfilm frame.	F	F
R4	User Interface	The system shall provide capabilities that support a user interface function. The user interface includes all aspects of the system that may be used directly by an individual.		
R401	Interactive and intuitive user interface	The user interface must be interactive and intuitive. It must provide for ease of training, use, and understanding. Major user interface objectives are to minimize key strokes, be self protecting (i.e., the system should not accept an invalid entry by an individual user), and self validating (wherever possible).	P	F
R402	Menus, prompting aids, & context-sensitive help	The system shall provide menus, prompting aids, indicate subsequent mandatory or optional steps, and provide context-sensitive help.	F	F
R403	Multiple levels of user interface (Novice, etc.)	Different levels of user interface shall be provided. A novice-user interface mode shall be provided with additional prompting aids to assist a user in understanding and using the features and capabilities of the system.	F	F
R404	Consistent design to user interface	The user interface shall be consistent (i.e., <F1> is always Help). It shall provide a standard dialogue that has predictable characteristics for all types of user requests, tasks, and system functions. The consistency of the user interface must include the design of input and output formats, both in general layout and in the use of uniform terminology, icons, and abbreviations.	P	F

F=full support, P=partial support, N=no support, blank=unknown

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R405	Unambiguous user interface - clear messages	The user interface shall be unambiguous. Each display, output report, and all alarm or error messages must indicate what is represented and why it is being presented. Each user action must have an unequivocal response such as a positive acknowledgement for successfully completed actions or for confirming processing is occurring for requests that will take a long time to complete. The user must be able to clearly distinguish between system-oriented messages and responses to user requests.	P	F
R406	Flexible and convenient user interface	The user interface shall be flexible and convenient. It must support reasonable combinations or sequences of user actions required to accomplish a task such that the individual does not have to resort to awkward interactions or to manual record keeping.	F	F
R407	Ability to cancel incomplete requests	The system must allow an individual to edit and/or cancel incomplete requests and to cancel or curtail unwanted output.	F	F
R408	Ability to go forward (or back) n pages of text	The ability to go forward or backward "n" pages or screens in the ASCII text display shall be provided.	P	F
R5	Inputs	The system shall provide capabilities that support inputs to the system.		
501	Rapid entry of abstract and full text material	The system will provide for rapid and timely availability of document abstracts of the full text and for rapid and timely capturing of the full text content of selected documents.	P	F
R502	Deletion of documents requires special access	After a specified period for verification, any errors identified may not be corrected by revising the original document. Rather, the submitter must submit a corrected version. Both the header for the revised document and the original document shall note that two versions are in the system. Deletion of documents shall require special access.	F	F
R503	Rapid entry of header information	The system must provide for the rapid availability of input of header field data.	F	F
R504	Pre-editing of full text must be simple	The process for entering full text content into the system must be simple. Involved and time consuming conversion programs or routines are not acceptable.	P	F

F=full support, P=partial support, N=no support, blank=unknown

Requirement	Title	Definition	System	
			NUDOCS	PROTOTYPE
R6	Outputs	The system shall provide the capabilities that support the generation of system outputs.		
R601	Output of statistics, text, or combinations	The output from a query shall be either statistics (e.g., the number of documents containing a searched word), text, the value of header fields, or some combination of these elements.	P	F
R602	Support for system administration reports	The system shall support the output of system administration information such as period statistics on terminal activity (e.g., usage time, number of searches).	F	
R603	Ability to send output to local printer	The system shall support the capability to direct output to an attached or network local printer.	N	F
R604	Ability to format user-defined reports	The system shall support the capability of user defined formats for reports.	N	F
R605	Ability to cut and paste into other documents	The system shall support the ability of a user to "cut" portions of the result set ASCII text, and "paste" the cut portion into another document. This capability presumes the ability to save the cut portion of the text for import to a word processing software package or to "hot-key" between the document management system and a word processing application.	P	

F=full support, P=partial support, N=no support, blank=unknown